

Master of Science in Data Science

Dissertation

Analysis of the effect of course structure and pattern of usage on the efficacy of online/blended courses



By

Andries Cornelius Oosthuizen

Supervised by

Leanne Scott

Department of Statistical Science

University of Cape Town

12 December 2023

The copyright of this thesis vests in the author. No quotation from it or information derived from it is to be published without full acknowledgement of the source. The thesis is to be used for private study or non-commercial research purposes only.

Published by the University of Cape Town (UCT) in terms of the non-exclusive license granted to UCT by the author.

Plagiarism Declaration

I know the meaning of plagiarism and declare that all the work in the document, save for that which is properly acknowledged, is my own. This thesis/dissertation has been submitted to the Turnitin module (or equivalent similarity and originality checking software), and I confirm that my supervisor has seen my report and any concerns revealed by such have been resolved with my supervisor.

Signed by candidate

Signature of Author: at **Cape Town** on **12 December 2023**

Acknowledgements

I extend my thanks to the following people for their invaluable assistance in the production of this dissertation:

- My supervisor, Assoc Prof. Leanne Scott, for her help and guidance through this whole process and her unwavering patience.
- Stephen Marquard, for his advice and help.
- Carina Oosthuizen, my lovely wife, for her support through the writing and re-writing, talking through all the different approaches I wanted to try, understanding when I worked longer hours due to this study, and her patience for getting through this.

Abstract

Using the Sakai Learning Management System (LMS), this dissertation investigates the structure of course sites in blended and online courses at the University of Cape Town. Then, it evaluates the student interactions that this facilitates. The data selected focused on undergraduate courses in 2019. The student interactions with the tool selected for each site are compared to tool categories that indicate a good academic outcome. The analysis was structured to use four popular unsupervised learning algorithms (K-means, PAM, AGNES, and DIANA) on data sets that included the enrolled users and the tools accessible to students. The `clValid` package method was used to choose the optimal algorithm and cluster sizes.

The findings show that most sites used the default tool selection, with almost half the courses adding outside tools and linking in lecture recordings. Sites with less enrolled students were shown to include more diffuse tools, which allow for more creative pedagogy. The majority of student interactions were for course development and delivery, followed by grading and assessment.

Finally, most students utilised the LMS and accessed a high percentage of tools in each category. However, the analysis had certain limitations about the events tracked by the system and assumed a one-sided perspective as only the student interaction with the LMS was considered.

Table of Contents

Plagiarism Declaration	i
Acknowledgements.....	ii
Abstract.....	iii
Table of Contents.....	iv
List of Tables	vii
List of Figures	ix
Glossary.....	xii
1 Introduction	1
1.1 Aim.....	2
1.2 Objectives	2
1.3 Limitations.....	2
1.4 Assumptions	3
1.5 Consent.....	3
1.6 Dissertation Structure	3
2 Literature Review	4
2.1 Comparison of Learning Management Systems (LMS)	4
2.2 LMS Importance in Higher Education.....	7
2.3 Overview of the Tools Used in LMSs	8
2.4 Importance of Student Interaction with LMS Tools	9
2.5 What is the Limitation of an LMS	10
2.6 Blended Learning.....	10
2.7 Conclusion	11
3 Methodology.....	12
3.1 Defining the Period of Study	12
3.2 Learning Efficacy.....	13
3.3 Unsupervised Learning Algorithms.....	13
3.3.1.1 K-Means.....	14
3.3.1.2 K-Medoids Clustering – Partitioning Around Medoids (PAM)	14
3.3.1.3 Hierarchical Clustering (AGNES & DIANA).....	15
4 Data Analysis.....	16
4.1 Extracting Data - Exploratory Data Analysis (EDA)	16
4.1.1 Introduction	16
4.1.2 What Defines a Course Site?	16
4.1.3 Exploring the Data Structures to Determine the Structure of a Site	17

4.1.4	Initial Data Export.....	17
4.1.5	Providers and Realms.....	18
4.1.6	Exclusion of Duplicate, Test, or Sandbox Sites.....	29
4.1.7	Understanding Site Participants.....	33
4.1.8	Sakai Database Structure – Site and Tool Tables.....	46
4.1.9	Advanced Data Export.....	49
4.1.10	Accessing Outside Data Sources.....	50
4.1.11	Exporting Event Data.....	53
4.1.12	Summary.....	55
4.2	Processing Data.....	56
4.2.1	Data Set and Processing Steps.....	56
4.2.2	Course Participants.....	57
4.2.2.1	K-Means.....	58
4.2.2.2	K-Medoids Clustering – Partitioning Around Medoids (PAM).....	61
4.2.2.3	Hierarchical Clustering (AGNES & DIANA).....	62
4.2.2.4	Comparing Clustering Algorithms (Introducing clValid).....	64
4.2.2.5	Comparing Clusters to Event Data.....	68
4.2.3	Tools Used in Course Sites.....	70
4.2.3.1	Tools Visible to Students - Clustering.....	75
4.2.3.2	Tools Visible to Students - Comparing Clusters to Event Data.....	78
4.2.3.3	Categories of Tools Visible to Students – Clustering.....	82
4.2.3.4	Categories of Tools Visible to Students - Comparing Clusters to Event Data.....	85
4.2.4	Summary.....	85
5	Discussion.....	86
5.1	Introduction.....	86
5.2	Enrolled Students and Events.....	86
5.3	Tools Used in Course Sites.....	86
5.4	Student Interactions.....	88
6	Conclusion.....	89
6.1	Overview.....	89
6.2	Research Objectives.....	89
6.3	Future Work.....	91
	References.....	92
	Literature.....	92
	R.....	98
	Appendix.....	100

Appendix – Sakai Tools	100
Communications and Collaboration	101
Grading and Assessment.....	101
Course Development and Delivery	102
Powerful System Administration	102
Course Management.....	102
User Tools	102
System Administrator Tools.....	103
Community Contributed Tools.....	103
External App Integrations	104
Appendix – LMS Emotional Footprint.....	105
Appendix – LMS Category Overview.....	106
Appendix – Vendor Capability Summary	107
Appendix – 2019 Academic Year	108
Appendix – Participant Data Set - DIANA	109
Appendix – Correlation Matrix for Tools Visible to Students	110
Appendix – Tools Visible to Students - PCA	112

List of Tables

Table 1: SoftwareReviews - LMS Category overview of product performance (Best Education Learning Management (ELMS) Systems 2023, 2023) (Complete figure/table from source is shown on page 106)	6
Table 2: LMS selection for feature comparison (Blackboard (Anthology), 2023; Brightspace (D2L), 2023; Canvas (Instructure), 2023; Moodle LMS, 2023; Sakai LMS, 2023)	6
Table 3: LMS Feature Capabilities Comparison	7
Table 4: Sakai tool category summary	8
Table 5: Number of 'Course' and 'Project' sites in Sakai by date, provider, or term	16
Table 6: Roles available in a course and project site.	17
Table 7: Provider Structure – Course Code	19
Table 8: Provider Structure – Program Code	19
Table 9: Provider Structure Example	19
Table 10: Total number of courses for each academic career code.....	20
Table 11: Faculty codes with total number of courses, undergraduate and medical degrees.....	22
Table 12: Course site linked to multiple faculties.....	23
Table 13: Course sites without faculty or department.....	24
Table 14: Course sites for MEDS+PGDP or MEDS+NDGU.....	25
Table 15: MEDS academic career code sites (Top 10) ordered by number of providers.	25
Table 16: Combination of UGRD academic career code (Top 10) ordered by number of providers.....	27
Table 17: Course sites with NA term	27
Table 18: Course sites where the term is not 2019.....	28
Table 19: Course site with low total tool count, exception sites.....	29
Table 20: Sites filtered by "test" or "sandbox" in the title.	30
Table 21: Test sites for FTX3045S.....	30
Table 22: Course sites with the same title.....	31
Table 23: Duplicate titles that are published with a valid provider.....	32
Table 24: Comparison of duplicate sites showing the number of students and events.....	32
Table 25: Course site where Lecturer participants are 0.....	34
Table 26: Course site participants ordered by number of Site Owner descending (Top 10).....	35
Table 27: Course site participants ordered by number of Lecturers descending (Top 10)	35
Table 28: Course site participants ordered by number of Support Staff and Tutors (Top 10)	37
Table 29: Course site participants ordered by number of "Other" (Top 10).....	37
Table 30: Course site participants ordered by Number of Students (Top 10).....	38
Table 31: Number of Sites per Number of Students (Bottom 10)	40
Table 32: Course site with 0 students (15)	40
Table 33: Course site with 1 or 2 Students (7/15)	41
Table 34: Course site order by Student to Lecturer ratio (Top 10).....	42
Table 35: The lowest percentage of students and lecturers per course site.....	45
Table 36: Sakai total events and sessions per year (2017, 2018, 2019,2020)	53
Table 37: Total participants for 7 roles in the course data set (997).....	57
Table 38: Total participants for 4 combined roles in course data set (997)	57
Table 39: Hopkins statistics for the participant's data set.....	58
Table 40: K-means (k=2) min:mean:max (standard deviation) of participants in summarised roles	60
Table 41: PAM (k=2) min:mean:max (standard deviation) of participants in summarised roles	62
Table 42: AGNES (k=2) min:mean:max (standard deviation) of participants in summarised roles.....	64
Table 43: Participant cluster comparison (Internal) – Table.....	65
Table 44: Participant (clValid) Stability.....	68
Table 45: Cluster min:mean:max (sd) for K-means (k=2) for participants.....	69
Table 46: Cluster min:mean:max (sd) for AGNES (k=2) for participants.....	70
Table 47: Event, student, and tool details per category.....	72
Table 48: min:mean:max for events, students and tools visible to students in course sites	72
Table 49: Hopkins statistics for all tools and tools visible to students' data set.	74

Table 50: Course site sorted by total number of tools visible to students (Total Tools - lowest 10).	74
Table 51: Course site sorted by total number of tools visible to students (Total Tools - top 10)	74
Table 52: Tools visible to students' cluster comparison (Internal) – Table	75
Table 53: Cluster min:mean:max (sd) for K-means and AGNES (k=2) for tools visible to students.....	77
Table 54: K-means (k=2) sites close to the centre of clusters for tools visible to students.....	78
Table 55: AGNES (k=2) sites close to the centre of clusters using tools visible to students.....	78
Table 56: Course participants for clusters in K-means (k=2) for tools visible to students.....	80
Table 57: K-means (k=2) event to tool category comparison for tools visible to students	80
Table 58: Course participants for clusters in AGNES (k=2) for tools visible to students	82
Table 59: AGNES (k=2) event to tool category comparison for tools visible to students	82
Table 60: Categories of tools visible to students' cluster comparison (Internal) – optimal	82
Table 61: AGNES (k=2) event to tool for categories of tools visible to students.....	84
Table 62: AGNES (k=2) sites in the centre of cluster for categorised tools visible to students	84
Table 63: Course participants for clusters in AGNES (k=2) for categorised tools visible to students.....	85
Table 64: AGNES (k=2) event to tool category comparison for categorised tools visible to students	85
Table 65: Tools Used at Crestview Online (Table 1 - Kerr, 2011)	100
Table 66: Tools Used at Morris Online Academy (Table 2 - Kerr, 2011).....	100
Table 67: Tools Used in MRVC History Course (Table 3 - Kerr, 2011).....	100
Table 68: UCT 2019 Academic Year	108
Table 69: Subset of positive and negative values in the Correlation Matrix for Tools Visible to Students.	110
Table 70: Tools visible to students – PCA – Eigenvalues and Cumulative Variance	112

List of Figures

Figure 1: LMS Market Share for US & Canadian Higher Ed Institutions (2021) (Hill, 2022).....	4
Figure 2: SoftwareReviews - Data Quadrant for Education Learning Management Systems (Best Education Learning Management (ELMS) Systems 2023, 2023)	5
Figure 3: Relationship between site, tool, and property.....	18
Figure 4: Relationship between site, tool, property, provider, and role tables.	18
Figure 5: All course codes grouped per faculty - 2019 (SIS).	20
Figure 6: Number of course codes per academic career code grouped by faculty.	21
Figure 7: Only undergraduate course codes grouped per faculty - 2019 (SIS).	22
Figure 8: Number of course sites per faculty.....	22
Figure 9: Number of course sites per academic career code.	23
Figure 10: Number of course sites by academic career code filtered on undergrad and medical courses.....	23
Figure 11: Number of sites with no faculty or career code grouped by provider.	24
Figure 12: Number of course sites per academic career code filtered on MEDS.	24
Figure 13: Frequency histogram of provider links in MEDS academic career code.	25
Figure 14: Number of course sites for UGRD academic career code.	26
Figure 15: Frequency histogram of provider links in UGRD academic career code.	26
Figure 16: Number of course sites per term.....	27
Figure 17: Number of course sites per faculty before and after constraints have been applied.	29
Figure 18: Number of course sites per department grouped by faculty.	33
Figure 19: Histogram of the number of Lecturers and Site Owners in 2019 course sites	34
Figure 20: Histogram of the number of Support Staff and Tutors in 2019 course sites	36
Figure 21: Histogram of the number of Support Staff and Tutors (combined) in 2019 course sites	36
Figure 22: Histogram of the Other Roles (Reviewers, Observers, and Librarian) in 2019 Course Sites	37
Figure 23: Histogram of the Number of Students in 2019 Course Sites	38
Figure 24: Course site overview of "Tests STA1000S,2019" - indicating primary usage.	38
Figure 25: Course site overview of "Discover Commerce 2019" - indicating primary usages.	39
Figure 26: Course site overview of "CSC1015/17 Challenge,2019"	39
Figure 27: Histogram of the number of students in 2019 course sites with 50 or fewer students	40
Figure 28: Histogram of the Number of Students in 2019 Course Sites (with more than 3 students).	41
Figure 29: Histogram of the Number of Students in 2019 Course Sites	42
Figure 30: Ratio of Student to Lecturer and Students to Support by Faculty.....	43
Figure 31: Full-time equivalent (FTE) enrolments per academic staff member across the institution.	43
Figure 32: Number of Students to Lecturer in participants grouped by Faculty.	44
Figure 33: Percentage of Students and Lecturers in participants grouped by Faculty	45
Figure 34: Relationship between the site table and all the tool tables.	46
Figure 35: Number of discrete mime types, number of files in resources, and size of the files in resources.	48
Figure 36: Total Number of Lesson Types	49
Figure 37: Number of tools in All the course sites.....	50
Figure 38: Number of tool instances in course sites grouped by visibility to students.	51
Figure 39: Number of Opencast LTI instances in course sites	51
Figure 40: Opencast database structure.	52
Figure 41: Number of recordings for Opencast Series	52
Figure 42: Opencast recording view events by week from 2017-2020	53
Figure 43: Number of user interactions per week for 2017-2020	54
Figure 44: Number of user interactions per week on the filtered course sites for 2019.	54
Figure 45: Correlation matrix of course participants.....	57
Figure 46: Visual assessment of cluster tendency (VAT) of summarised roles	58
Figure 47: K-means – Determine the optimal number of clusters using (gap, wss, silhouette) for participants .	59
Figure 48: K-means (k=2) size of participant clusters	60
Figure 49: K-means (k=2) cluster silhouette of participants.....	60
Figure 50: K-means clustering of participants (k=2)	60

Figure 51: K-means (k=2) mean of roles	61
Figure 52: K-means (k=3) mean of roles	61
Figure 53: K-means (k=3) size of participant clusters	61
Figure 54: K-means (k=3) cluster silhouette of participants.....	61
Figure 55: PAM – Determine the optimal number of clusters using (gap, wss, silhouette) for participants	61
Figure 56: PAM (k=2) number of participants in clusters	62
Figure 57: PAM (k=2) cluster silhouette of participants	62
Figure 58: AGNES – Determine the optimal number of clusters using (gap, wss, silhouette) for participants	62
Figure 59: AGNES (k=2) size of participant clusters	63
Figure 60: AGNES (k=2) cluster silhouette of participants.....	63
Figure 61: AGNES (k=2) Dendrogram.....	63
Figure 62: AGNES (k=2) mean of roles.....	63
Figure 63: DIANA – Determine the optimal number of clusters using (gap, wss, silhouette) for participants	64
Figure 64: Participant cluster comparison (Internal) – Graphs.....	66
Figure 65: Participant cluster comparison (Stability) – Graph.....	67
Figure 66: Students to the total event in course sites	68
Figure 67: Students to the average event per student in course sites	69
Figure 68: K-means, AGNES, DIANA cluster sizes for participants (k=2).....	69
Figure 69: Tools visible to Lecturers and Students in the filtered course sites.	70
Figure 70: Number of students to the tool choices grouped by category	71
Figure 71: Number of events per tool grouped by tool category	72
Figure 72: Students' correlation to tools	73
Figure 73: Distribution of the tools in course sites (all tools or only visible to students).	73
Figure 74: Tools visible to students' cluster comparison (Internal) – Graphs	76
Figure 75: DIANA (k=2) cluster size for tools visible to students	76
Figure 76: DIANA (k=3) cluster silhouette for tools visible to students.....	76
Figure 77: K-means (k=2) cluster size for tools visible to students.....	77
Figure 78: K-means (k=3) cluster silhouette for tools visible to students	77
Figure 79: AGNES (k=2) cluster size for tools visible to students	77
Figure 80: AGNES (k=3) cluster silhouette for tools visible to students	77
Figure 81: Count of event types in event data	79
Figure 82: Number of tools on a site to total events.....	79
Figure 83: Categories of tools visible to students' cluster comparison (Internal) – Graphs.....	83
Figure 84: AGNES – Determine the optimal number of clusters using (gap, wss, silhouette) for categorised tools visible to students.....	83
Figure 85: AGNES (k=2) cluster size for categorised tools visible to students.....	84
Figure 86: AGNES (k=3) cluster silhouette for categorised tools visible to students.....	84
Figure 87: AGNES (k=2) dendrogram for categorised tools visible to students.....	84
Figure 88: Average percentage of active students and events per student (min:median:mean:max)	88
Figure 89: Software Reviews - Emotional Footprint for Education Learning Management System (SoftwareReviews, 2020).....	105
Figure 90: SoftwareReviews - LMS Category overview of product performance (Complete) (Best Education Learning Management (ELMS) Systems 2023, 2023)	106
Figure 91: SoftwareReviews - LMS Category overview of Vendor Capabilities (Best Education Learning Management (ELMS) Systems 2023, 2023)	107
Figure 92: DIANA (k=2) size of participant clusters	109
Figure 93: DIANA (k=2) cluster silhouette of participants	109
Figure 94: DIANA (k=3) size of participant clusters	109
Figure 95: DIANA (k=3) cluster silhouette of participants	109
Figure 96: DIANA (k=5) size of participant clusters	109
Figure 97: DIANA (k=5) cluster silhouette of participants	109
Figure 98: Correlation Matrix for Tools Visible to Students	110
Figure 99: Correlation of tool categories in “Tools Visible to Students”	111

Figure 100: Contribution of variables to the PCA components for tools visible to students 113
Figure 101: Biplot of principal components, including Cos2 contribution values for PCA 113
Figure 102: PCA – Cos2 of variables on the first 10 dimensions in tools visible to students..... 114

Glossary

ABBREVIATION	TERM	CONTEXT
AGNES	Hierarchical Clustering - Agglomerative Nesting	Data Science
API	Application Programming Interface	Coding
BI	Business Intelligence (SAP BI)	Environment
BO	Business Objects (SAP BO)	Environment
BUSN	Business Degree	University
CLARA	Clustering Large Applications	Data Science
CMS	Course Management System	Environment
COVID-19	Coronavirus Disease 2019	Environment
CSV	Comma Separated Values	Coding
DIANA	Hierarchical Clustering - Divisive Analysis	Data Science
DOCT	Doctoral Degree	University
EDA	Exploratory Data Analysis	Data Science
ELMS	Educational Learning Management System (See LMS)	Environment
HONS	Honours Degree	University
ICT	Information and Communications Technology	Environment
IETF	Internet Engineering Task Force	Environment
IT	Information Technology	Environment
LMS	Learning Management System	Environment
LPUP	Learning Platforms Update Project	Environment
LTI	Learning Tool Integration	Environment
MAST	Masters Degree	University
MEDS	Medical Degree	University
NDGLE	Next-Generation Digital Learning Environment	Environment
NDGP	Postgraduate Non-Degree	University
NDGU	Undergraduate Non-Degree	University
PAM	Partitioning Around Medoids	Data Science
PDOC	Post-Doctoral Degree	University
PGDP	Post Graduate Diploma	University
REST	Representational State Transfer	Coding
SAP	Systems Applications and Products in Data Processing (Company Name)	Environment
SIS	Student Information System	Environment
SOAP	Simple Object Access Protocol	Coding
SQL	Structured Query Language	Coding
UCT	University of Cape Town	Environment
UGRD	Undergraduate Degree	University
URL	Uniform Resource Locator	Coding
UWC	University of Western Cape	Environment
UX	User Experience	Environment
VAT	Visual Assessment for Tendency	Data Science
VLE	Virtual Learning Environments	Environment

1 Introduction

In the higher education learning environment, the Learning Management System (LMS) has become a de facto standard of delivering content, administrating courses, and facilitating communication and interaction between students and lecturers. The LMS market is a multi-billion-dollar industry valued at around \$18.7 billion (2022) and is projected to grow to \$43.6 billion by 2027 (*Learning Management System Market Size & Global Forecast, 2023*). In North America (US and Canada) higher education market, the primary system tends to be one of 5, namely Instructure's Canvas, Blackboard Learn, D2L Brightspace, Moodle, and Sakai (Hill, 2022; *LMS Data – Spring 2021 Updates, 2021*).

In terms of breadth of features and quality of features, Sakai ranks higher than Canvas, Moodle, and Blackboard (O'Brien L, 2020). Regarding the number of default available tools, not including Learning Tool Interoperability (LTI), Sakai provides the most extensive set. The open-source nature also allows access to the tools' internal structure, data storage and events. This access to the internal structure provides unfettered access to the internal storage of events and site structures used by lecturers to scaffold the teaching and learning experience.

The University of Cape Town has been using the Sakai, called Vula, since 2009 (the earliest course site), but only properly from around 2013. Sakai is a free community source system (community of academic institutions, commercial organisations, and individuals) and is distributed under the Educational Community License (a type of open-source license). The first version of Sakai was released on 27 October 2004 (Hodges et al., 2023), but only after version 2.0 (2006) was released did it start to gain mainstream interest (C. R. Severance, 2013) and is currently running at over 240 institutions worldwide (Aperio, 2023b).

This study investigates the different tools and modules instructors utilise in a course site in Sakai to deliver educational materials and facilitate communication, assessment, and grading. The selection of tools may vary depending on the subject, faculty, and personal preference. Moreover, students have unique usage patterns that may differ by subject and the site's design. Evaluating which tools are effective in facilitating the learning experience for students will benefit the further development of such tools and ultimately benefit the students, university, and broader open-source community.

1.1 Aim

Learning Management Systems (LMS) are extensively used in higher education institutions to support and extend face-to-face teaching and learning (blended learning). The Learning Management System (LMS) is a complex platform with various features and functionalities. Its architectural design supports the system's functions and performance behaviour. The layout, tools, and user experience (UX) inform the choices that lecturers make in scaffolding the student learning experience. The tools chosen in a course site allow interactions between learners and other learners, learners and their teachers, and learners and course content (Kerr, 2011; Moore, 1989).

This research will analyse the course structure and content available to students to look at the pattern of usage to identify groups of similarly structured courses. Utilising the course groups, the user interactions will be analysed to determine the efficacy of the course design. Efficacy in this regard will be evaluated on the amount of interaction with specific tools and content. The focus will be on blended and online courses, as the content and user interactions are all stored in the LMS.

There have been other studies conducted in the field of Learning Analytics to evaluate student engagement with the LMS to determine educational effectiveness and learning outcomes (Fisher & Parker, 2016; Le et al., 2022; Zhang et al., 2020), relating student interaction to academic achievements. The scope of these studies was generally limited to specific courses, departments, or the effectiveness of a new tool. This study aims to expand the scope of site selection and also focus on structural differences that influence student interactions, in so doing contributing to the expanding body of exploratory research in the area of Learning Analytics. The additional information may inform further studies on effective tool usage and improvements in LMSs.

1.2 Objectives

The objectives include:

Literature Review

1. Examine the Learning Management System (LMS) marketplace.
2. Investigate the importance of LMSs in higher education.
3. Research the impact of student interactions with tools on learning outcome results.

Data and Analysis

4. Collect data from Sakai (LMS) pertaining to course structure and student interactions.
5. Analyse the course structure and classify.
6. Compare student activity based on the classification of the course structure.

Evaluation and Conclusion

7. To recommend further research or development of tools.

1.3 Limitations

The study is limited by the data available in the LMS chosen by the University of Cape Town. The data includes course sites, tools, and interactions; the interactions are limited to what the developers of each tool deemed important to track and store in the interaction log. This inconsistency in the importance of events might skew the evaluation of tool interactions, as some tools might not have any interactions stored in the log. The pedagogical strategy for the course sites is unknown and is not used in the analysis phase. Furthermore, this study does not evaluate the efficacy of learning; effectiveness is determined by interactions with the tools, not academic achievement.

1.4 Assumptions

- 1) It is assumed that all students included in this study have sufficient access to the University of Cape Town's Learning Management System (Sakai named Vula).
- 2) It is assumed that all students are sufficiently proficient in using the LMS.
- 3) It is assumed that lecturers use specific tools to achieve a learning outcome for students.
- 4) It is assumed that the tools used in the course site will track the progress of students by generating events that will be stored in the event table in the database.
- 5) It is assumed that the number of events generated for each tool is generally the same for an activity performed by the student.
- 6) The events tracked are for formal learning activities and exclude interactions outside the scope of the Learning Management System.

1.5 Consent

When students register at the university, they agree to allow the university to record their academic information using various information technology systems such as academic administration system, access control system, IT system and residence management system. The university regularly uses this data for administrative and research purposes.

At the time of this research, Andries Cornelius Oosthuizen was employed as a senior developer in the Learning Technologies Team in the Centre of Innovation in Learning and Teaching at the University of Cape Town. As such, access to the data forms an integral part of his job description.

1.6 Dissertation Structure

This study consists of six chapters, starting with the introduction and overview in Chapter 1.

Chapter 2 introduces the Learning Management System (LMS) and its current role in higher education and the general marketplace. This chapter compares the tools available in an LMS and how students interact with them and examine how they affect academic outcomes. Finally, the chapter concludes with a discussion on the limitations of LMSs and the definition of blended learning.

Chapter 3 describes the data selection criteria and introduces the unsupervised learning algorithms used in this study.

Chapter 4 covers the concept of a course site the extraction of the site structure from the complex Sakai database tables, closely followed by data preparation and cleaning. The chapter also continues to explain what a provider, realm, and participant is in the Sakai LMS context. The relationship between the tools available in the LMS and how they are mapped in the database is shown. After this, the event data and outside sources are introduced. The last section starts with the constraints applied to the course site data set.

The second part of Chapter 4 concerns the processing of the data by employing unsupervised learning algorithms discussed in Chapter 3. Next, the tools used in the course sites are analysed similarly and compared with the event data collected earlier.

Chapter 5 presents the results and findings of this study, followed by a conclusion in Chapter 6 that summarises the results and identifies potential opportunities for future research.

2 Literature Review

A Learning Management System (LMS) is a software system that provides tools to support a learning/teaching process (Caminero & Hernández, 2013). These systems are also known as Virtual Learning Environments (VLE) or Course Management Systems (CMS) (Asiri et al., 2012; Falvo & Johnson, 2007). Whatever one may call them, they are used pervasively in higher education institutions (Cucu, 2014); estimates of colleges and universities running an LMS are almost always near 99 per cent (Brown et al., 2015; Mehta & Kalyvaki, 2017). This is expanded on by (Rhode et al., 2017) in which they state that the use of LMSs in higher education is widespread, with 99% of institutions having an LMS in place and 85% of faculty and 83% of students using them (Rhode et al., 2017).

2.1 Comparison of Learning Management Systems (LMS)

As an educational institution, choosing which LMS to use is an involved and lengthy process. Each LMS has their strengths and weaknesses, and there have been numerous studies to find, define, and classify features to evaluate and decide on which one to use (Asiri et al., 2012; Cavus & Alhih, 2014; Falvo & Johnson, 2007; Reyes et al., 2009). LMSs are generally grouped into commercial or non-commercial, with closed, open, or partially open architecture (Pireva et al., 2015). Non-commercial open (as in open source) are LMSs such as Sakai, Moodle, and Ilias. Then there are the commercial partially open, like Canvas (*Open-Core Model - Wikipedia, 2023*). The last category comprises commercial closed architecture, such as Brightspace and Blackboard (Caminero & Hernández, 2013). This section will compare a few of the most popular LMSs based on market size and influence.

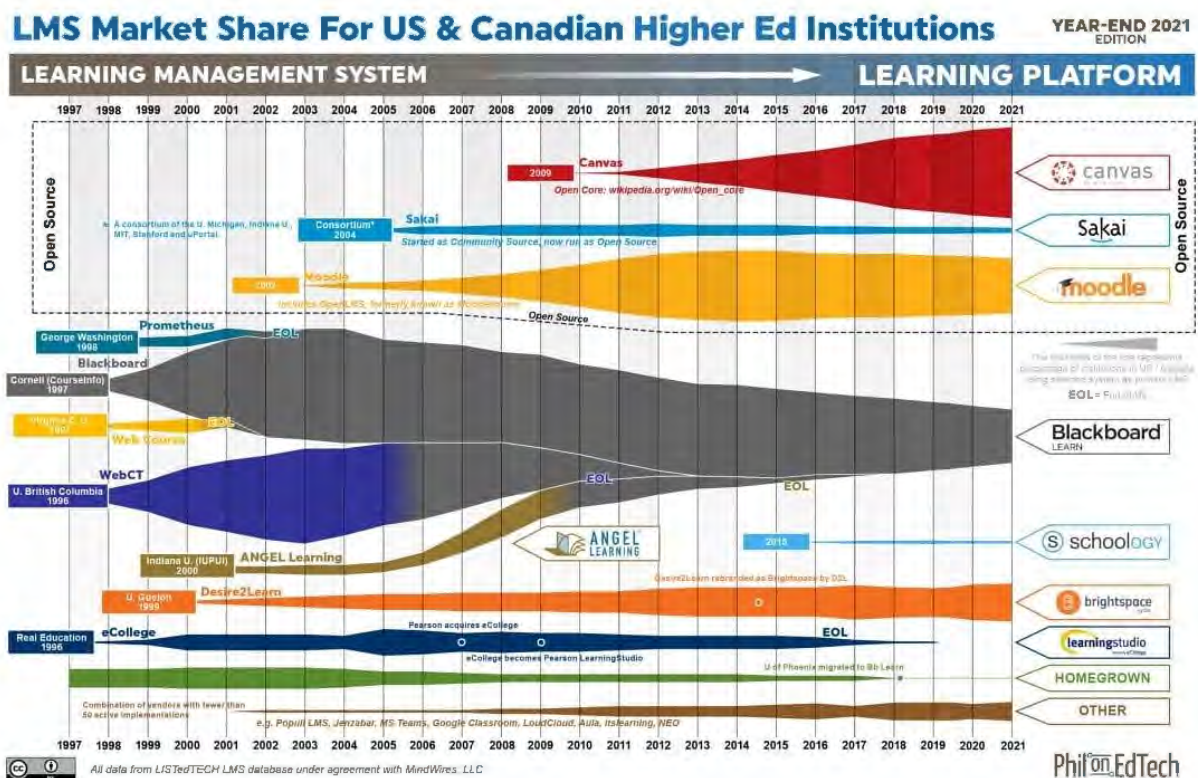


Figure 1: LMS Market Share for US & Canadian Higher Ed Institutions (2021) (Hill, 2022)

Figure 1 (above) from Phil-on-Tech (now Phil Hill and Associates) published on "State of Higher Ed LMS Market for US and Canada: Year-End 2021" in Feb 2022 indicates the relative market share of various LMSs in the United States and Canada. It is clear that the LMS market has been around for a

couple of decades and has experienced significant changes. The newest data at the end of 2021 shows Canvas leads with 34% of US & Canadian higher ed institutions, followed by Moodle at 21%, Blackboard at 20%, and D2L at 14% (Hill, 2022). Another company that specialises in evaluating software, SoftwareReviews, provides Figure 2 (below) showing the evaluation and ranks of LMSs based on feedback from information technology and business professionals. The y-axis shows the product features and satisfaction, which is derived from the likelihood to recommend the software combined with the satisfaction of product features, plotted against vendor capabilities (training, customer support, product roadmap, trustworthiness, respect, fairness, etc.). Note the axes are dynamically scaled.

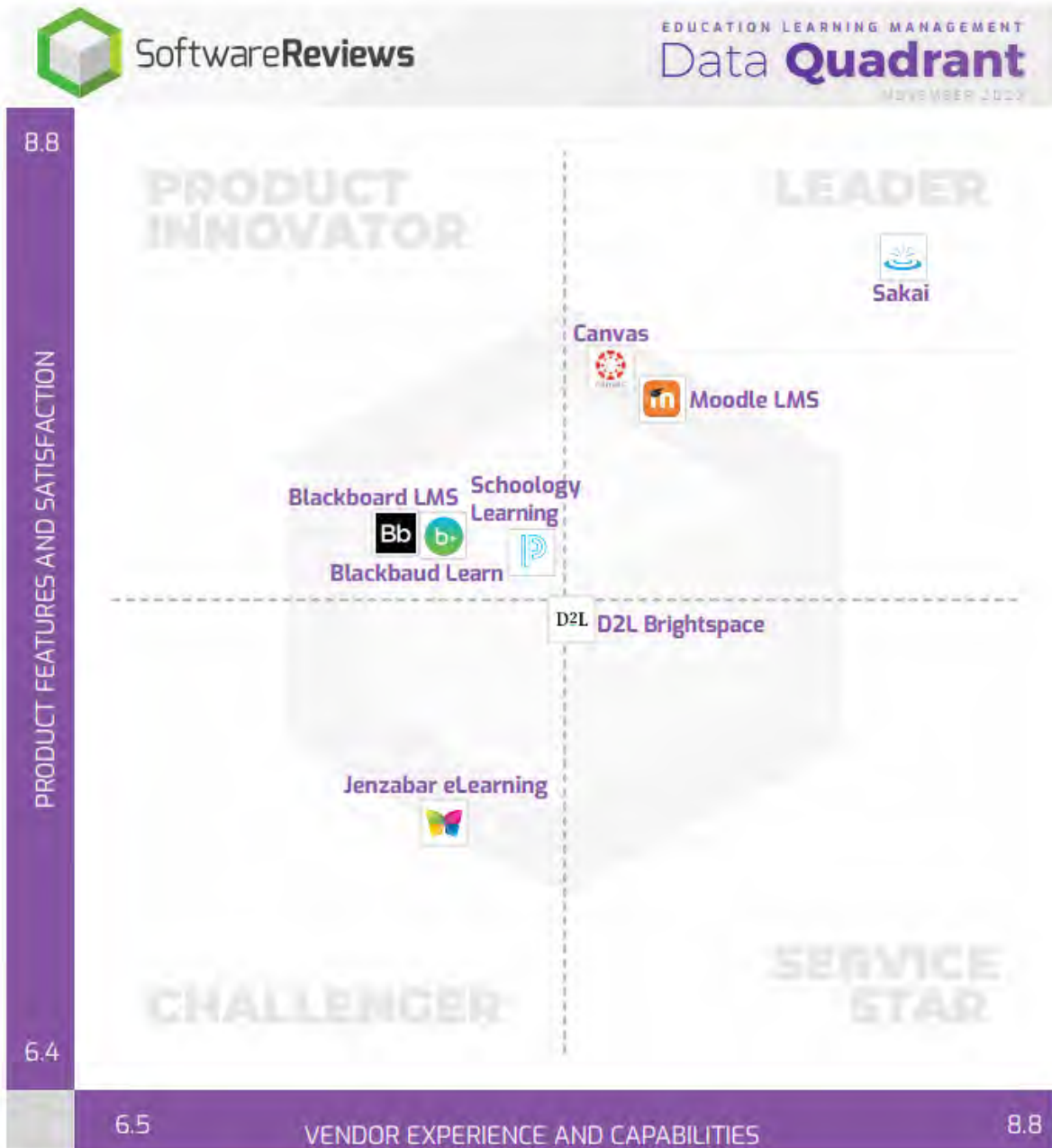


Figure 2: SoftwareReviews - Data Quadrant for Education Learning Management Systems (Best Education Learning Management (ELMS) Systems 2023, 2023)

Rank	Product	Composite Score (Out of 10)	Net Emotional Footprint (-100 to 100)	Vendor Capabilities	Product Features	Likeliness to Recommend	Number of Reviews
1	Sakai LMS	8.5	+90	81%	83%	87%	98
2	Moodle LMS	8.0	+81	76%	79%	82%	206
3	Canvas	7.9	+77	77%	79%	84%	197
4	Schoology Learning	7.5	+76	73%	74%	78%	106
5	D2L Brightspace	7.5	+76	76%	77%	72%	51
6	Blackbaud Learn	7.4	+69	75%	74%	79%	42
7	Blackboard LMS	7.4	+66	75%	77%	76%	305
8	Jenzabar eLearning	7.0	+73	71%	75%	62%	34

Table 1: SoftwareReviews - LMS Category overview of product performance (Best Education Learning Management (ELMS) Systems 2023, 2023) (Complete figure/table from source is shown on page 106)

Table 1 (above) shows the summary of product performance for the top 8 LMSs based on their composite score, which is a combination of their Net Emotional Footprint (which measures user emotional response ratings of the vendor, e.g., trustworthy, respectful, fair), Vendor Capabilities, Product Features, and Likeliness to Recommend (the original table which includes Net Emotional Footprint Distribution can be found in the Appendix – LMS Category Overview).

The emotional footprint might be based on each LMS’s aesthetics and tool design, but they are generally similarly matched on features (Pireva et al., 2015). These features or tools are used to create and manage content for learners, do evaluations or assessments with grading and communicate (synchronous or asynchronous) with students (Caminero & Hernández, 2013; Falvo & Johnson, 2007; Kerr, 2011). Similarly, the tools that students identify as important to their learning experience (especially communication with their instructor) (Kerr, 2011) can generally be found in successful LMSs. Looking at the features of the top five market share leaders and high-ranking LMS, there are a few commonalities between them.






				
Sakai LMS	Moodle LMS	Canvas	D2L Brightspace	Blackboard LMS Platform

Table 2: LMS selection for feature comparison (Blackboard (Anthology), 2023; Brightspace (D2L), 2023; Canvas (Instructure), 2023; Moodle LMS, 2023; Sakai LMS, 2023)

This comparison table (Table 2) is based on (Caminero & Hernández, 2013; Cavus & Zabadi, 2014; Firdyiwiek, 1999; Kerr, 2011; Pireva et al., 2015; Reyes et al., 2009) with additional details obtained from the documentation of each of the LMSs (*Blackboard Tools*, 2023; *Brightspace Tools*, 2023; *Canvas Tools*, 2023; *Moodle Tools*, 2023; *Sakai 23 User Guide*, 2023).

The check (✓) indicates this is a built-in feature, and the cross (✗) indicates that the feature is not provided at all. The tilde (~) indicates that the feature or tool is provided as an additional add-on, plugin, or tool integration - directly or with Learning Tool Integration (LTI).

Feature Capabilities		Sakai	Moodle	Canvas	Brightspace	Blackboard
Course Management	Calendar	✓	✓	✓	✓	✓
	Structure Content	✓	✓	✓	✓	✓
	Student Statistics	✓	✓	✓	✓	✓
Synchronous Communication	Chat	✓	✓	✓	✓	✓
	Video	~	✓	~	~	✓
	White Board	~	✓	~	~	✓
Asynchronous Communication	Forum	✓	✓	✓	✓	✓
	Email	✓	✓	✓	✓	✓
Testing and assessment	Grading	✓	✓	✓	✓	✓
	Assignment	✓	✓	✓	✓	✓
	Quizzes	✓	✓	✓	✓	✓
	Surveys	✓	✓	✓	~	✓
Student Authoring	Wiki	✓	✓	~	~	✓
	Portfolio	~	✓	✓	✓	✓
Scaffolding Tools	Content Organising	✓	✓	✓	✓	✓
	Annotations	~	✓	✓	✓	✓
Accessibility		✓	✓	✓	✓	✓
Software Ecosystem	Web Architecture	Tomcat	Apache	Tomcat	-	Tomcat
	Database	MySQL, Oracle	MySQL, PostgreSQL, MSSQL, Oracle	PostgreSQL, MSSQL, MySQL, Oracle	-	MSSQL, Oracle
	Marketplace	Open Source ECL-2.0	Open-Source GPL-3.0	Open-Source AGPL-3.0	Commercial	Commercial
Supported Standards	SCORM	✓	✓	✓	✓	✓
	IMS LTI	LTI 1.3	LTI 1.3	LTI 1.3	LTI 1.3	LTI 1.3
	xAPI / Caliper	✓✓	✓✓	✓✓	✓✓	✓✓
Customisable	Environment	✓	✓	✓	✓	✓
	Profile	✓	✓	✓	✓	✓

Table 3: LMS Feature Capabilities Comparison

Table 3 compares the features and tools for five platforms, namely Sakai, Moodle, Canvas, Brightspace, and Blackboard; firstly, based on the tool categories and then based on the software ecosystem, standards, and customisability of the LMS (overall and user profile). Although there are a few features that do not match up exactly, these LMSs provide all the necessary features and capabilities for higher education institutions to support the learning project.

2.2 LMS Importance in Higher Education

Regardless of the LMS that an institution is using, they are seen not only as a technology platform for supporting the traditional learning process but also for networking, blended, and distance learning. (Pireva et al., 2015). The trend in modern education is highly technological dependent, and this has redefined the teaching-learning process. LMSs have a positive impact on education (Cavus & Alhih, 2014). Further studies in US institutions indicate that nearly all higher education students use an

LMS and that students consistently identify the LMS as among the most important instructional technologies for their academic success (Brooks & Pomerantz, 2016; Rhode et al., 2017). Not just in the US, the flexible capabilities of learning management systems (LMSs) have led to an increasing acceptance of web-enabled e-learning tools among institutions of higher education globally (Czerniewicz et al., 2007; Mlitwa et al., 2009).

LMSs have also been found to have a transformative effect on traditional teaching and learning in adult and higher education institutions. They influence the method of acquiring higher education, facilitate transformative learning, and are considered game-changers in the field (Reid & Reid, 2019). It can be shown that faculty also believes that the LMS contributes to effective classroom management and dissemination of knowledge. It also helps with their flexibility in teaching and grading students' assessments on a weekly basis (Reid & Reid, 2019).

LMSs in higher education institutions are an indispensable tool in the modern teaching-learning process, hence their adoption (Czerniewicz et al., 2007; Mlitwa et al., 2009).

2.3 Overview of the Tools Used in LMSs

This study will focus primarily on Sakai as the LMS (reasons and history will be discussed in later sections). As one of the suites of open-source LMSs, Sakai is more popular in higher education institutions. It offers a basic set of learning tools and has the capacity to easily add or adapt new tools (Guo, 2013). Its other advantages include a user-friendly interface, particularly in terms of communication tools (Cavus & Zabadi, 2014). The usage of Sakai as a learning tool extends beyond traditional classroom settings. It has been adopted by universities to create online environments for distance education students and highlights the versatility of Sakai's tools, which can facilitate learning in various contexts (Amponsah et al., 2019).

Sakai's tools can be apportioned into a few categories (*Sakai Features*, 2023). Table 4 shows the categories and some examples of each – the complete list can be found in the "Appendix – Sakai Tools".

Category	Tools
Communication and Collaboration (Communication)	Announcements, Chat, Commons, Forums, etc.
Grading and Assessment (Assessment)	Gradebook, Assignments, Rubrics, Test & Quizzes etc.
Course Development and Delivery (Content)	Lessons, News, Resources, Syllabus etc.
Course Management (Management)	Statistics, Roster, Polls, User Tools, etc.
Community Contributed Tools (Various)	Attendance, BlogWow, Evaluation System, etc.
User Tools (System)	Profile, Membership, Preferences, etc.
System Administrator Tools (System)	User, Server, Memory, etc.
External Integrations (Integrations)	Examiity, eXplorance Blue, Opencast, Tsugi, Turnitin etc.

Table 4: Sakai tool category summary

2.4 Importance of Student Interaction with LMS Tools

The effectiveness of an LMS and its impact on student learning and performance has been a topic of interest in several studies with various approaches to defining user interaction and obtaining data related to these interactions.

In the study of LMS adoption in Nigeria by (Yakubu et al., 2020), their research shows that using an LMS has the potential to improve instruction and learning outcomes (Yakubu et al., 2020). Highlighted in the study is the factor of student acceptance and adoption of the LMS, which is influenced by the perceived value from people that students regard as important (Yakubu et al., 2020). This ties into the finding from (Rhode et al., 2017) which indicate that student use is both driven and limited by faculty use and to the findings from (Fisher & Parker, 2016), suggesting that the efficacy of LMS systems is based on whether they are well perceived by the students using them.

Learning Analytics, extracted from the Moodle LMS logs, can be used to predict a student's performance, as in the study by (Mwalumbwe & Mtebe, 2017); the results support the hypothesis that interactions with the LMS have a positive impact on student performance. It was found that tools like Forums, Peer Review and Interaction, and Exercises (Test & Quizzes) had the most impact. In contrast, the number of downloads, login frequency, and time spent in the LMS had no impact on students' performance (Mwalumbwe & Mtebe, 2017). This can help institutions identify students who may be at risk of poor performance and provide timely interventions.

In a similar study by (Zhang et al., 2020), the student activity in the Moodle logs were extracted and used in a correlation analysis to determine the impact of student's activity on the final assessment. It was found that interaction with the course site, particularly opening files (higher frequency of opening files), positively correlate with a higher grade; it also found that opening a file event in the logs correlate positively with activity in a forum discussion, which supports the findings in the previous study.

In a study by (Le et al., 2022), it was observed that high-achieving students accessed the LMS resources more frequently than mid or low-achieving students. It's interesting to note that in this study, peer interaction was among the lowest average number of clicks per student, likely due to the fact that peer collaboration in the LMS is often not obligatory (Le et al., 2022).

The study by (Sáiz-Manzanares et al., 2019) investigated the use of LMS with hypermedia Smart Tutoring Systems with personalised student feedback and found that it enhanced student engagement and understanding, leading to better learning outcomes.

Focusing just on online assignments (Ahmed & Mesonovich, 2019) found that there was no doubt that harnessing technology can enhance learning, whether it is inside or outside of the classroom. The key, however, is how effective the learning tools are used in the classroom, the manner of which requires a great deal of work, time, and patience.

In a study on the students' behaviour in the Sakai LMS, the students who performed excellently in learning spent more time in the Lessons tool and contributed actively to the discussion forums (Wan et al., 2018).

In a similar study by (Avcı & Ergün, 2019), the students' online activities were analysed to determine their effect on engagement. In that particular paper, it was found that "students with high participation had both high engagement and high performance. It can, therefore, be put forward that high levels of student participation can lead to effective learning."

In Fisher & Parker (2016), it was found that in half of the courses analysed, there existed a significant positive relationship between both measures of online learning resource utilisation frequency (i.e., course site visits and online learning resource downloads) and achieved course grade. This was found to significantly change across the various course types, course levels of study and course sessions. The study used student interactions in Sakai for all the students registered in a particular faculty (Faculty of Commerce at the University of Cape Town), which was exported to the SAP Business Intelligence platform.

2.5 What is the Limitation of an LMS

As with all software systems, there are limitations to what the system can achieve based on architecture, development decisions and user requirements. In the case of eLearning systems (LMS, VLE, CMS), these limitations can be defined in the restrictions of the tools to facilitate the learning process through the delivery of learning materials (content, assessment, communication, etc.) and the peer and social interactions it encourages or not.

Significant advancements have been made in terms of technology, architecture, functionality, and content presentation. However, there are still key functionalities that are missing, such as content and course personalisation, better content structuring, customisable syllabus and learning paths, ontology adaptation, etc. (Pireva et al., 2015). More research is needed to understand the use of tools to support higher-order thinking and how multiple tools interact to create deep learning experiences for students in online learning situations (Kerr, 2011). Additionally the impact of LMS use on pedagogical practices (Holmes & Prieto-Rodriguez, 2018) is still an area of ongoing research.

It is also important to note that not all learners take full advantage of the benefits that the LMS might provide (Al-Rahmi et al., 2015). This might be due to the user's computer literacy (ICT skills) and/or time restrictions (Pireva et al., 2015).

There is a view that the restrictions on tools and content in an LMS should be unbundled to facilitate more agile and open ways of accessing learning components. The existing LMS is seen as too limiting, and the proposition is to move to what is often referred to as the "next-generation digital learning environment" (NGDLE) to support more personalised and flexible learning experiences. Rather than being a single system, the NGDLE would encompass a "confederation of IT systems and application components that adhere to common standards...that would enable diversity while fostering coherence" (EDUCAUSE Learning Initiative, 2015; Rhode et al., 2017).

2.6 Blended Learning

There are two influential papers which define blended learning. Firstly (Garrison & Kanuka, 2004) defines blended learning as "the thoughtful integration of classroom face-to-face learning experiences with online learning experiences." Secondly (Graham, 2006) defines blended learning as follows: "Blended learning systems combine face-to-face instruction with computer-mediated instruction". These two definitions differ on some points, as noted in (Hrastinski, 2019), but it shows that blended learning has become an umbrella term for any combination of face-to-face and online learning. (Dhiman & Gera, 2021) highlights the fact that in blended learning, the LMS aspect of the student's learning environment lets them become an active participant in the teaching and learning process to enhance their academic achievement. In the context of this study, the course sites used in Sakai are used as supplementary learning content for the face-to-face classes and even capture those classes as Lecture Recordings, which students can then view at their leisure.

2.7 Conclusion

Overall, these studies suggest that the use of an LMS at a higher educational institution does have the potential to positively impact the student learning experience. The efficacy depends on the use of the LMS by the instructors, their ability to tailor learning experiences, facilitate peer interactions, and the incorporation of interactive and multimedia elements to improve learning outcomes for students. The scaffolding for the learning process is blended, meaning that there are face-to-face classes with learning content available in the LMS.

The literature does highlight that there are limitations in mapping the direct interactions of learners to their academic achievements; this is mainly due to the learning process being an internal process, not just owing to their interactions with the LMS. Although most of the studies have limitations on the number of courses, learner interactions, and tool selection, it was generally found that using an LMS had a positive outcome.

This study aims to supplement previous research by including a more comprehensive range of tools and course sites. Earlier studies, as identified in the literature review, were limited to specific courses and tools or a closely related set of courses. In contrast, this study includes all available tools in the Sakai LMS and all the undergraduate course sites for 2019. This is achievable due to the direct access to the course site structure and student-generated events for each tool. By broadening the scope of the tools and course sites, this study aims to review the tools selected and their usage for undergraduate courses in the Sakai LMS.

3 Methodology

3.1 Defining the Period of Study

The data set used for this study was selected based on day-to-day operations at the University of Cape Town (UCT), where blended learning has been implemented across most faculties and departments. The time frame for the data set selection was defined by two major events that took place during the time of this study.

The first event was the #FeesMustFall movement, which began in the first few months of 2015 and was repeated in 2017 and 2018 (Cini, 2019). The movement emerged in response to proposed tuition fee hikes at universities, which sparked nationwide student protests (Cini, 2019). This tied into a movement that was more specific to UCT, the #RhodesMustFall, where students demanded the statue of Cecil John Rhodes be removed from their campus; as a consequence, the students at the University of Western Cape claim that UCT sparked the #FeesMustFall movement (Langa et al., 2017). Overall, the protests on various campuses caused multiple universities to suspend teaching and learning for periods of time until such time as campuses could be re-opened and made safe from interruption to classes. At UCT, this interruption in everyday teaching is visible in the event summary graphs (Figure 43).

The next major impact on the learning project was in the early months of 2020; Coronavirus Disease 2019 (COVID-19) emerged as a global pandemic, which caused the South African Government to declare a national disaster on Sunday, 15 March 2020 and to implement a nationwide lockdown that started on Thursday, 26 March which lasted up until Thursday, 16 April 2020 (21 Days). This lockdown had a significant impact on higher education institutions in South Africa. It limited face-to-face classes and the movement of students to campus at the start of the year. This also necessitated the development of alternative ways to ensure the learning project could continue (Landa et al., 2021). One of the strategies implemented was the introduction of virtual classes (Mhlanga & Moloi, 2020). UCT took the threat of infection extremely seriously and instituted a transition to fully online learning, which only returned to some normality for some faculties in 2021. This affected the choice of tools for sites and their overall usage, as can be seen for 2020 events on the total number of interactions graph (Figure 43) later in this document.

Both of these unprecedented events, which in their own way disrupted the learning project, affected the manner in which teaching could proceed at the university and, in doing so, influenced the choice of tools and student behaviour.

As part of Vision 2030 (*Vision 2030*, 2023) at the University of Cape Town, the Learning Platforms Update Project (LPUP) (*Learning Platforms Update Project*, 2023) was announced in May 2021. The project aims to choose a suite of online learning platforms to meet the needs of teaching and learning for the next 5 to 10 years. After a rigorous evaluation of the current LMS (Sakai) and its competitors, the committee decided to switch over to D2L Brightspace in early 2022 (*The University of Cape Town Selects D2L To Enhance and Future Proof Its Digital Learning Infrastructure*, 2022). The Brightspace LMS was branded as Amathuba (*Amathuba*, 2023) and, per the project plan, will replace Sakai (Vula) LMS by 2025. The project required migrating 50 initial courses across faculties and departments for the 2nd semester of 2022, affecting 55 teaching staff and over 2,000 students in the pilot phase. Currently, there have been 467 complete migrations for 2022 course sites and 823 for 2023. This number is expected to increase during the last months of 2023 and into the first semester of 2024, as the project plan calls for all undergraduate courses to be taught in Amathuba by 2024.

Taking all of the events, incidents, and projects into account for the last couple of years, the most normalised year, in terms of blended learning, could be seen as 2019.

The data for 2019 is extracted from the Sakai Database and Web Service APIs (Application Programming Interface). The information from the extracted data shows a course site's structure, its associated tools, and users' interactions with the learning materials. (Burgos, 2019) suggest the information captured around user interactions could help identify specific character traits about users, which can be used to create predictive models or, in other cases, describe uniqueness. Using the tool choices in conjunction with the user interactions could allow educational stakeholders, such as lecturers, support staff, and university management, the knowledge to leverage LMS site construction to improve the framework for student learning.

3.2 Learning Efficacy

Learning efficacy refers to an individual's belief in their ability to learn and perform tasks successfully this is closely related to self-efficacy, which can be “defined as the belief individuals have in their own capacities, in this case, the capacity to learn”. The studies referred to in “Antecedents of employees' involvement in work-related learning: A systematic review” show that individuals with higher levels of self-efficacy are more likely to engage in learning opportunities and experience greater learning success, both in formal and informal learning settings (Kyndt & Baert, 2013).

Learning efficacy is also closely related to student interactions in an LMS. The social context of learning and interactions within the course sites (lecturers and fellow students) can significantly impact students' sense of valuation of goals, efficacy, and the nature of the task (Roderick & Engel, 2007).

The literature supports the view that students' behaviour using the Sakai LMS can be used to evaluate their performance (Wan et al., 2018) and that high-performing students spend more time using the Lessons tool and activities in discussion forums (Wan et al., 2018). Another supporting study on Sakai data (Fisher & Parker, 2016) found that there “existed a significant positive relationship between both measures of online learning resource utilisation frequency (i.e., course site visits and online learning resource downloads) and achieved course grade”. The efficacy depends on the student's confidence in their abilities to perform specific tasks within the LMS, such as navigating to the learning material, completing assessments, and overall engagement with the tools in the site. This is impacted by the lecturer's design of the learning materials in the course site.

In this study, efficacy will be defined as the effectiveness of the LMS in facilitating student interaction and engagement in the various tools available on a course site.

3.3 Unsupervised Learning Algorithms

The 2019 data set contains 2,509 course sites. There are 47,486 users enrolled as students and 9,802 in various other roles, which include lecturers (2,915), support staff (1,175), tutors (2,018), and observers (1,354). Out of all the users enrolled in the course sites, only 46,511 have a valid UCT student number, which suggests that some users enrolled as students are using staff (2,485), third-party (222) or guest accounts (2,378). The students enrolled in the course sites generated 167,843,604 events as stored in the event table of Sakai.

The sites in the data set include undergraduate and postgraduate courses. The scope of the study limits the selection to sites used for undergraduate blended learning courses. The course sites are

then clustered around participants and tool selection to find a representative site in each cluster that can be used to evaluate the user interactions with the chosen tools. This approach might provide insight into the group of tools selected for a course site and how this impacts students' level of activity. The literature indicates that "high participation had both high engagement and high performance" and "that high levels of student participation can lead to effective learning" (Avcı & Ergün, 2019).

In order to select the best representation of sites, various clustering methods will be assessed to determine which one is the most suitable for clustering the sites around participants, and tool choices.

The clustering algorithms described in this paper are K-means, Partitioning Around Medoids (PAM), and Hierarchical Clustering (bottom-up [AGNES] and top-down [DIANA]). K-means is one of the simplest and most popular used unsupervised learning algorithms (Sinaga & Yang, 2020; Helm, 2021). It divides the observations into K number of clusters based on their similarities. This is similar to the paper (Burgos, 2019), where the K-means algorithm was used to cluster students' cognitive styles based on survey findings. In (Avcı & Ergün, 2019), K-means is used in conjunction with Hierarchical clustering to group students based on their behaviour. PAM models the observation similarly to K-means, but instead of using the nearest mean of the cluster, it uses an observation chosen from the data set to be the medoid/prototype of the cluster (Schubert & Rousseeuw, 2019). As an alternative to PAM, Clustering Large Applications (CLARA) was considered as it is an extension of PAM, but it is more suited for a large number of objects (several thousand observations) (Kaufman & Rousseeuw, 1990; Sinaga & Yang, 2020). This data set did not warrant the use of such an algorithm based on the low number of observations. Hierarchical Clustering Analysis is an unsupervised learning algorithm that classifies related observations across high-dimensional data (Avcı & Ergün, 2019; Lee et al., 2016).

3.3.1.1 K-Means

The K-means clustering algorithm defines clusters where the total within-cluster variation is minimised. To do this, the number of clusters is specified (named k), and then a random selection of k objects is taken from the data set and set as initial cluster centres or means. Then iteratively, the following steps are performed (Davidson, 2022; Maechler et al., 2022):

1. Assign each observation to its closest centroid, using the Euclidean distance between the observation and the centroid.
2. For each cluster (k), update the cluster centroid by using the new mean values of all the observations in that cluster.
3. Calculate the total within sum of squares and check exit conditions – cluster assignments stop changing, or the maximum of iterations is reached; then loop back to 1.

3.3.1.2 K-Medoids Clustering – Partitioning Around Medoids (PAM)

The K-medoids clustering, more widely known as Partitioning Around Medoids (PAM), is related to K-means clustering for partitioning a data set into a number of (k) clusters (Davidson, 2022). The centres of the clusters are called medoids. The medoid is an observation which is the most centrally located in a cluster of observations, meaning that the average dissimilarity between all of the other members in the clusters has the lowest value (minimal). The medoids are representative of the cluster and can be used in analysing the clusters. In contrast, with K-means, the cluster's centre is a calculated mean point of all the observations in a cluster, not an actual observation.

This algorithm is also a bit more robust than K-means as it is less sensitive to outliers and noise. As with K-means, K-medoids require a user to specify k . The PAM algorithm consists of the following steps (Davidson, 2022; Maechler et al., 2022):

1. Randomly select k observations to become the medoids.
2. Calculate the dissimilarity matrix (Manhattan or Euclidean).
3. Assign every observation to the closest medoid.
4. Review each cluster. If any observation in that cluster decreases the average dissimilarity coefficient the most for that cluster, then select it as the new medoid for the cluster. If no medoids have changed – end the loop.

3.3.1.3 Hierarchical Clustering (AGNES & DIANA)

Hierarchical clustering is a widely used unsupervised machine learning algorithm that groups similar objects together based on their characteristics or distances between them. The main objective is to create a hierarchy of clusters by either merging (bottom-up) or splitting (top-down) existing clusters iteratively until a stopping criterion is reached. The algorithm is suitable for a range of data types and mixed data sets. Hierarchical clustering is quite computationally intensive and, as such, is better suited for small to medium-sized data sets. Unlike the previous clustering methods, the hierarchical clustering algorithm does not require a pre-specified number of clusters (Kassambara, 2023e).

3.3.1.3.1 Agglomerative Nesting Hierarchical Clustering (AGNES)

The most common type of hierarchical clustering is the AGglomerative NESTing (AGNES) approach, where the observations are grouped based on similarity. This bottom-up approach divides all the observations into “leaf” nodes, single element clusters, and then, on each step of the algorithm, the two most similar clusters are combined (distance calculation). This is done iteratively until just one node, the root, is left. The result is a tree-based representation of the objects, named a dendrogram (Davidson, 2022; Kassambara, 2023a), visually showing the order of cluster mergers. This approach is used more commonly because it is computationally efficient.

3.3.1.3.2 Divisive Analysis Hierarchical Clustering (DIANA)

The top-down hierarchical clustering approach is called DIvisive ANALysis Hierarchical Clustering (DIANA). In this method, all the observations start in a single cluster (root) and then recursively split into smaller clusters until each observation ends up in its own cluster (leaf). Divisive clustering is not frequently used because of its significant computational complexity (Davidson, 2022; Kassambara, 2023d).

4 Data Analysis

4.1 Extracting Data - Exploratory Data Analysis (EDA)

4.1.1 Introduction

As already stated, the sites selected for this study will include undergraduate blended learning courses, which will be used to investigate their structure and draw conclusions about engagement. To this end, it is necessary to identify and exclude sites which are not primarily teaching and learning focused. This section explains the concept of a sites in Sakai (4.1.2), the data structures associated with them (4.1.3) and their relationship with tools (4.1.8). It will also explore the various ways that the data can be extracted (4.1.4) and which one is more suitable for this study (4.1.9). The different ways that Faculties, Departments and Lecturers at UCT use sites are discussed in 4.1.5 and 4.1.6.

4.1.2 What Defines a Course Site?

The first page displayed when accessing the Sakai URL (vula.uct.ac.za) is called the “Welcome” page (which incidentally is a “site” in Sakai called !gateway). The “Welcome” page displays some helpful links that users might find useful and the login forms to authenticate and access their learning environment. After the login, the user is presented with their homepage, an instance copy of the !home template site, which they can customise to their requirements. The course sites the user is enrolled in are accessible from this page, either in the site menu bar or the “View All Sites” panel. Accessing a course page allows users to complete tasks and access content, for example, completing an assignment or sending an announcement.

An example of a course site would be the site called “History of Economic Thought”. This course is taught in the Department “School of Economics” in the Faculty of Commerce, which then contains the learning content, assignments, and communication channels for that course.

In Sakai, sites can also be created to manage other activities like projects, conferences, clubs, or orientation-type sites. They have access to the same kind of tools, but their focus is mostly on communication in a community.

	BY CREATED YEAR (2019)	BY PROVIDER	BY YEAR
Project	2,659	-	-
Course	2,552	2 247	2 509

Table 5: Number of ‘Course’ and ‘Project’ sites in Sakai by date, provider, or term

Table 5 shows the two types of sites available in Sakai - Project and Course. Most project sites are not associated with a year, as they can run for multiple years or for a short period of time (i.e., a Conference). The second row shows the number of course sites created in 2019 - this likely includes sites destined for teaching in 2020. The next column indicates the number of sites where the provider link contains the 2019 term value. The last column indicates the number of course sites where the site property indicates that they are for teaching in 2019.

A provider is a university code associated with a course or program taught in a specific year; this code can be linked to a course site to provide user enrolment information. Note that Sakai can link multiple providers to a course site, which is rare in LMSs. On the creation of a course site, the year can be selected, and this is stored as the “term” site property.

It is important to note that each LMS has its distinct security policies associated with roles in the system (this might differ slightly for each implementation of an LMS). They might be named

similarly, but the access and authority will differ. In *this* Sakai, different roles are created for the type of site, and the privileges for those roles differ. On creation, the role templates are copied from the “type” template (site.template.course, site.template.project) and applied to the site.

COURSE	PROJECT
Site Owner	Site Owner
Lecturer	Project co-ordinator
Student	Participant
Support Staff	Support Staff
Tutor	Group manager
Observer	Observer
Librarian	-

Table 6: Roles available in a course and project site.

Table 6 shows the roles that can be assigned to a site by “type”. The important roles in our analysis are the “Lecturer”, “Site Owner”, “Support Staff” (all of whom control the site participants and tools) and the “Student”. The activity data of students will be used in determining the efficacy of the chosen site tools and access to content.

4.1.3 Exploring the Data Structures to Determine the Structure of a Site

Sakia can use a variety of databases, namely, Oracle, MySQL, MS SQL, and MariaDB. In this instance, the database is MySQL (version 5.7). Data in relational databases like MySQL are arranged into tables, with relationships – hence the name, which contains rows of data describing each data entry or instance of that type (table) and columns containing the entry's properties. Using the SAKAI_SITE table as an example, the rows would contain the properties of a site, such as the title, creation date, user that created the site, a unique identifier (SITE_ID), and other values. Other tables are linked (in relation) to the site record by referring to the SITE_ID. The SITE_ID is a Global Unique Identifier (GUID), a 128-bit unique reference number used in computing that is highly unlikely to repeat when generated despite there being no central GUID authority to ensure uniqueness.

4.1.4 Initial Data Export

Sakai is an open-source project, which allows us the unique opportunity to access the database structure directly (other LMSs do not allow us this kind of direct access). To extract data on a course site, it is required to construct a query in Structured Query Language (SQL), which defines what tables to access and what properties to retrieve. The result of the constructed queries can then be exported as Comma-separated values files (CSV). The queries and export formats are standard practice in most studies of this format and increase the reproducibility of the classification process.

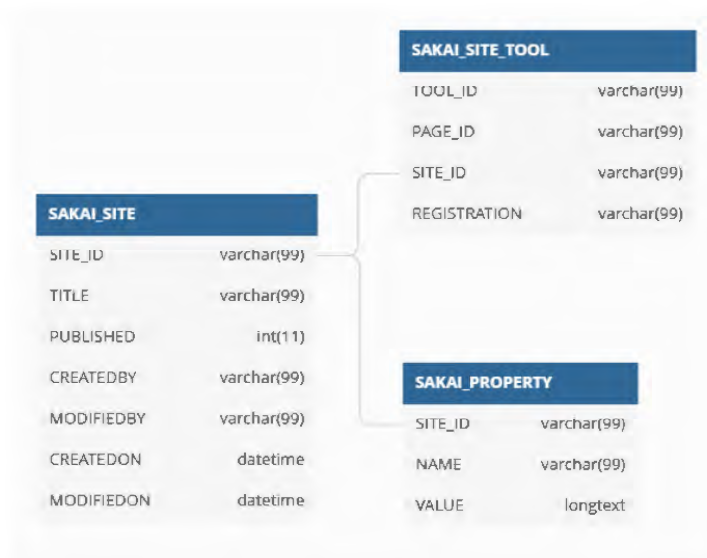


Figure 3: Relationship between site, tool, and property.

The initial export included only the course sites with the TERM property selected with the value of “2019” (SAKAI_PROPERTY name=”term” AND value=”2019”), the tools associated with that site and the values for SITE_ID, TITLE, published status, creation, and modification. The CSV files, which are easily inspected with Microsoft Excel, indicated that the site structure and details were correct and further processing and classification was then done in R.

4.1.5 Providers and Realms

The initial data export only contained information about the course site and associated tools. Additionally, the provider information gives us faculty, department, degree level, year, and user enrolments.

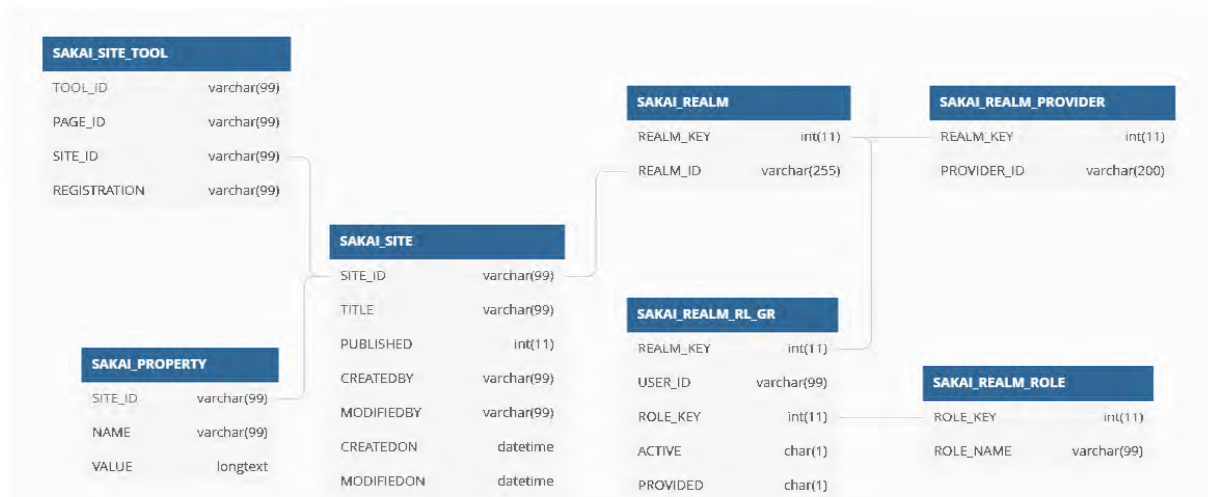


Figure 4: Relationship between site, tool, property, provider, and role tables.

The queries were improved to include the relational link between the site and the realm. A realm in Sakai is a combination of roles and permissions for a site, which means that the enrolment and provider link information can be extracted. In the table structure (Figure 4), the relationship between the site (SAKAI_SITE), realm (SAKAI_REALM), provider (SAKAI_REALM_PROVIDER), and the enrolment (SAKAI_REALM_RL_GR) tables can be seen. The SAKAI_REALM_RL_GR table contains the

user's unique ID and the relationship to SAKAI_REALM_ROLE, which defines what type of role the user is enrolled in.

The provider table contains the link from the university code to the site and, as such, also manages which users will be enrolled as a "Student". Sakai retrieves the enrolment information from the Student Information System (SIS) and processes the required updates and enrolments.

A provider is constructed as either a course code in this format:

AAAI_{nnn}B,YYYY

AAA	Is the three-alpha code of the department or unit administering the course or, in the case of a faculty-administered course, of the Dean's Office
I	Indicates the academic level of the course
nnn	Identifying numbers between 000 and 999
B	The 'year portion' indicator, or suffix, identifies the type of course and when it is offered (Explanation in attached appendix)
YYYY	The year of this course

Table 7: Provider Structure – Course Code

Or as a program code in this format:

FD_{nnn},YYYY

F	Single character denoting Faculty
D	Single character denoting the type of qualification (e.g., M – Masters)
nnn	Identifying numbers between 001 and 999
YYYY	The year of this course

Table 8: Provider Structure – Program Code

The course that was mentioned earlier (e.g., "History of Economic Thought") has the course code provider of "ECO3016F,2019", which breaks down as follows:

ECO	Course in the School of Economics
3	3 rd Year Course
016	ID of the course
F	1 st semester (half) course
2019	The year of the course is taught in

Table 9: Provider Structure Example

The same structure can be used to link multiple providers to a course site in the LMS; the course or program codes are concatenated with a "+" character; for example, the course "Secondary Piano, 2019" has the provider value of "MUZ1281H,2019+MUZ2281H,2019". This indicates that the course is provided by the South African College of Music (MUZ) and can be completed in the student's 1st or 2nd year; the course is taught over a year (H - Half course taught over a full year) in 2019.

Using the provider information, the analysis data set can be filtered to only include undergraduate courses taught in 2019. Sites with providers that include program codes generally overlap an entire department or faculty. From the initial findings, these courses are generally used for communication and not teaching and learning; as such, they will be excluded from the analysis data set.

Not all courses in the university are associated with Sakai sites; a data extract from the SIS for all courses in 2019 results in the following table.

Description	Academic Career Code	Total
Undergraduate	UGRD	1,848
Masters	MAST	1,477
Honours	HONS	634
Post-Graduate Diploma	PGDP	401
Medical	MEDS	364
Business Degree	BUSN	202
Doctorate	DOCT	173
Post-Doctoral	PDOC	54
*Postgrad Non-Degree	NDGP	38
*Undergrad Non-Degree	NDGU	12
		5,203

Table 10: Total number of courses for each academic career code.

* Used by occasional students and semester study abroad students. These students do not complete a degree; they only start certain courses for a term or semester.

The table above indicates that there are 5,203 course codes in use at the university for various academic career options. Note that there are 1,848 undergraduate course codes which fall under the scope of this study.

The courses can also be grouped by faculty:

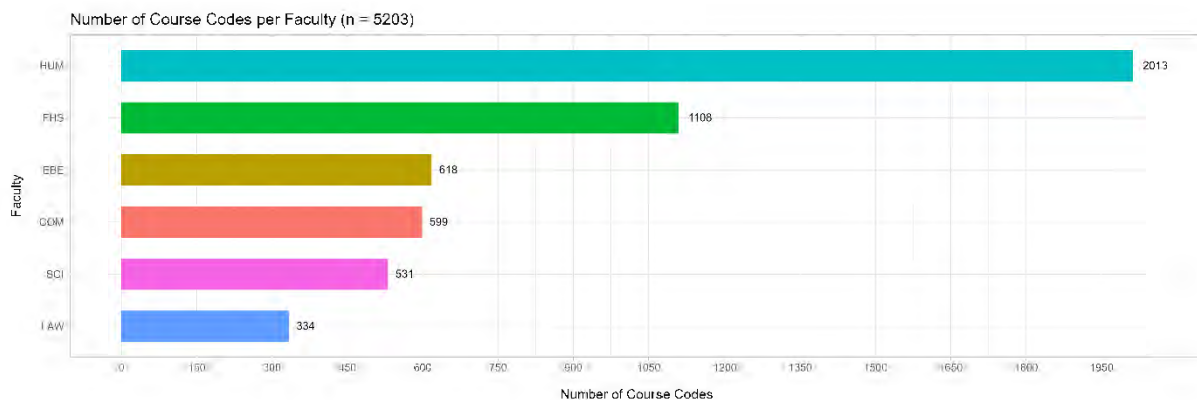


Figure 5: All course codes grouped per faculty - 2019 (SIS).

The graph (Figure 6) on the following page shows the academic career codes grouped by faculty.

Number of Course Codes per Academic Career Code Grouped by Faculty (n = 5203)

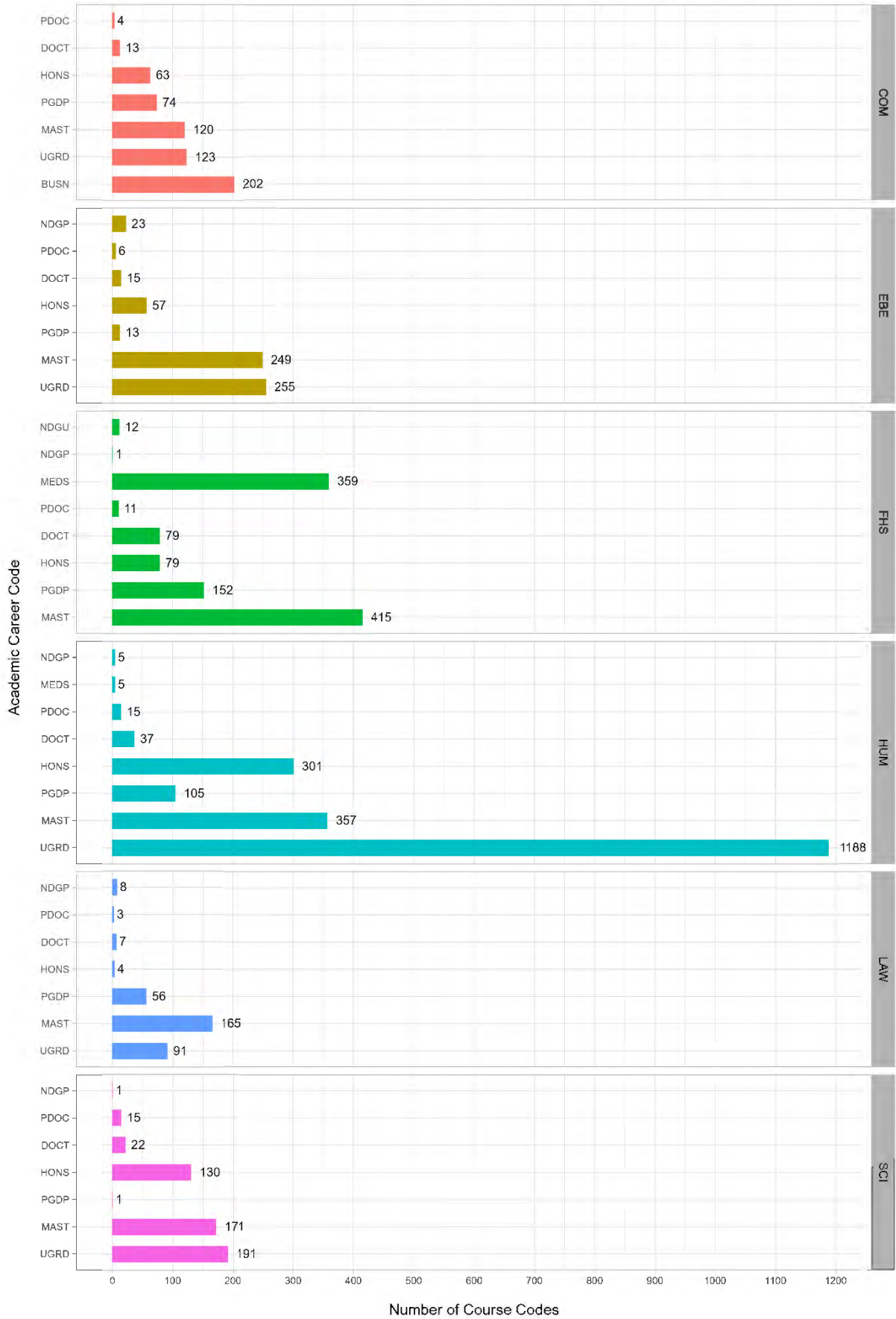


Figure 6: Number of course codes per academic career code grouped by faculty.

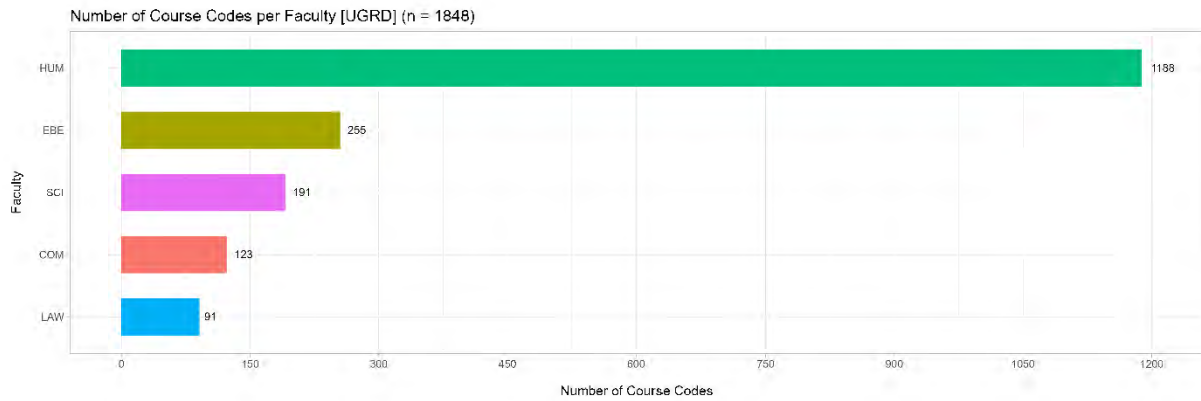


Figure 7: Only undergraduate course codes grouped per faculty - 2019 (SIS).

Figure 7 (above) indicate that the Faculty of Humanities has the most course codes and undergraduate courses. Note the absence of the Faculty of Health Sciences in the undergraduate graph; this faculty does not use the Undergraduate (UGRD) academic career code for their undergraduates (pre-doctor). They use the MEDS academic course code. Notably the course site data set might not match the number from the SIS source, as sites in Sakai can link multiple course codes to a single site (Table 11).

Faculty Code	Faculty	No. Course Codes	No. UGRD	No. MEDS
HUM	Faculty of Humanities	2013	1188	0
FHS	Faculty of Health Sciences	1108	0	359
EBE	Faculty of Engineering & Built Environment	618	255	0
COM	Faculty of Commerce	599	123	0
SCI	Faculty of Science	531	191	0
LAW	Faculty of Law	334	91	0
		5203	1848	359

Table 11: Faculty codes with total number of courses, undergraduate and medical degrees

In the next part of the analysis, the MEDS academic career code will be included, as undergraduate courses are associated with those course codes. Using the course site data and grouping them by faculty highlights a single course associated with multiple faculties and 152 sites with no faculty or department (empty provider link).

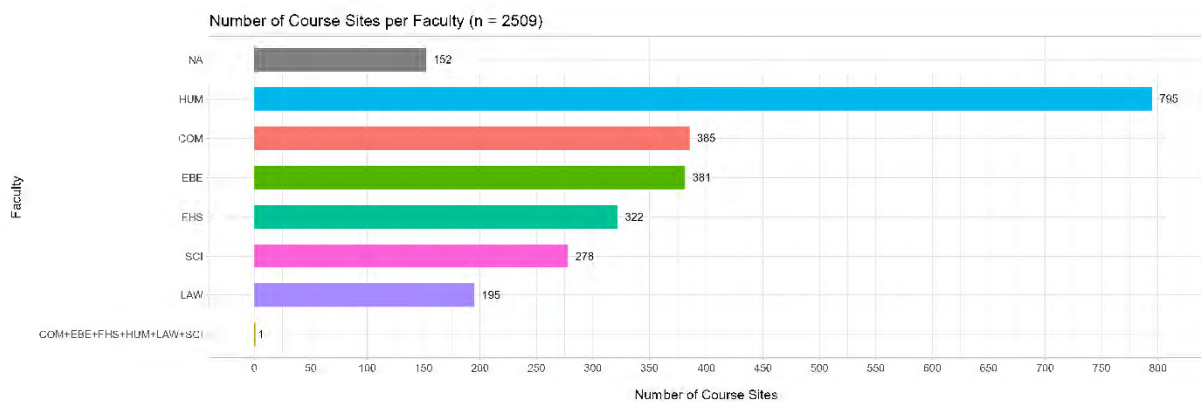


Figure 8: Number of course sites per faculty.

The course associated with multiple faculties is an information site for undergraduates to inform them about getting “Funding for Postgrad”. As such, it will be excluded from the subsequent data sets.

Title	Faculty	Academic Career	Term
Funding your Postgrad	COM+EBE+FHS+HUM+LAW+SCI	BUSN+DOCT+HONS+MAST+MEDS+UGRD	2019

Table 12: Course site linked to multiple faculties.

Mapping all the provider link combinations in Sakai course sites results in the following graph:

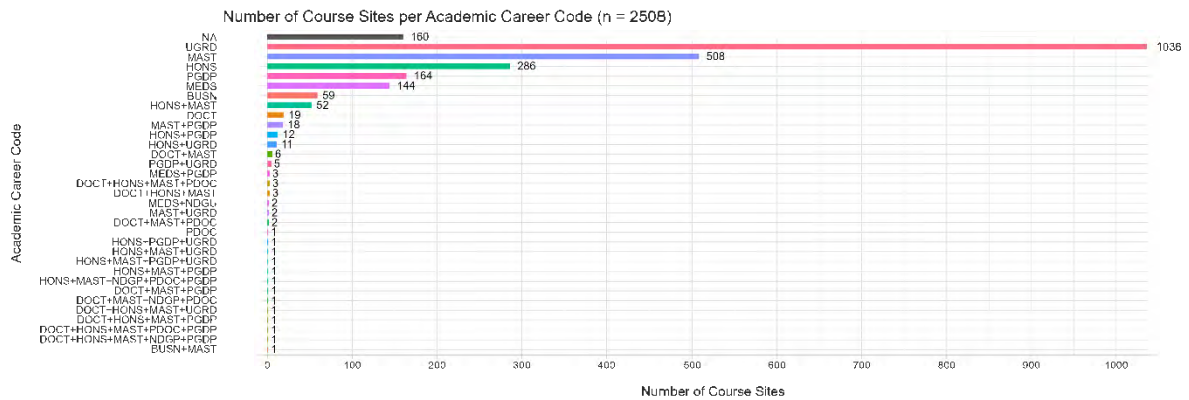


Figure 9: Number of course sites per academic career code.

It can be seen that there are various sites with combinations of academic career codes, filtering them on undergraduate and medical degree codes results in Figure 10.

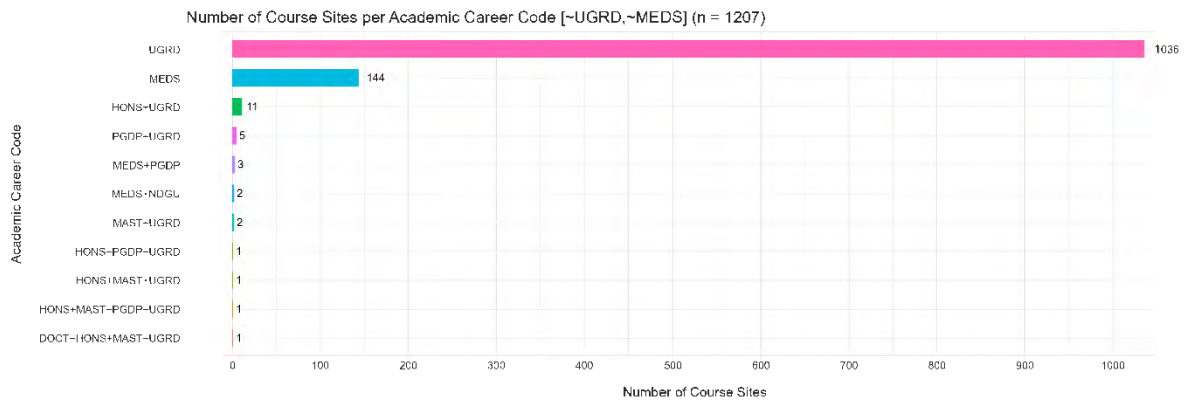


Figure 10: Number of course sites by academic career code filtered on undergrad and medical courses.

In Figure 9 (above) there are around 160 sites that do not have any provider information; thus, there is no academic career code. In that set, there are 12 “Turnitin” specific sites; these sites have been created by lecturers or admin staff as a way for students to submit assignments or, more likely, thesis/dissertation documents to the Turnitin plagiarism detection service.

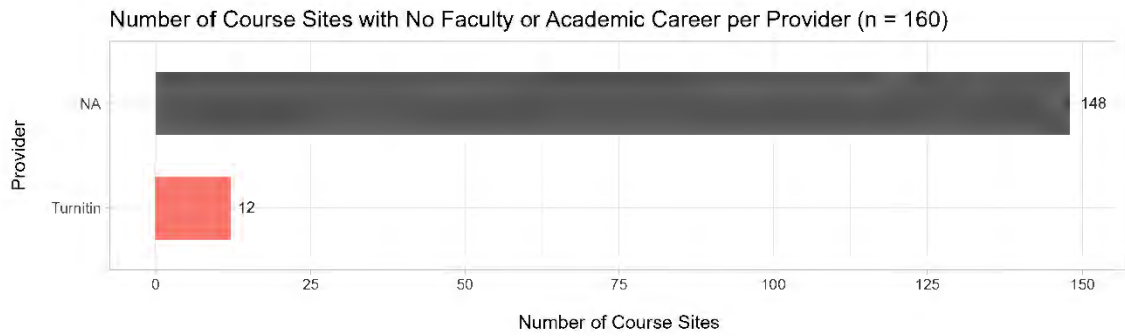


Figure 11: Number of sites with no faculty or career code grouped by provider.

This leaves 148 course sites without faculty, provider, or academic career codes. These sites (Table 13) contain various functions and do not fall into the scope of teaching and learning sites used in this analysis.

Title	Faculty	Academic Career	Term	Provider
Intro to PC March 2019	NA	NA	2019	NA
MCom in RMFM	NA	NA	2019	NA
Pilotsite	NA	NA	2019	NA
Mind Matters 2019	NA	NA	2019	NA
Special Studies Module - Music and Medicine	NA	NA	2019	NA
CILT Learning Designers	NA	NA	2019	NA
FYE Commerce - 2019	NA	NA	2019	NA
Testing Site	NA	NA	2019	NA
SAMLA-UCT Workshop 2	NA	NA	2019	NA
PG Dipl 2019/2020	NA	NA	2019	NA
SCIENCE FACULTY 2019	NA	NA	2019	NA
HCS 2019	NA	NA	2019	NA
EMBA20 2018-2020	NA	NA	2019	NA
PGDipHPE 2019: Course 4	NA	NA	2019	NA
Derm Short Course 2019	NA	NA	2019	NA
Test Board Sem 6, 2019	NA	NA	2019	NA
CEC1 unused draft 2019	NA	NA	2019	NA
Test Board Sem 3, 2019	NA	NA	2019	NA

Table 13: Course sites without faculty or department.

Figure 12 uses the information from Figure 10 as the base and focuses just on the MEDS academic career code and the combinations thereof.

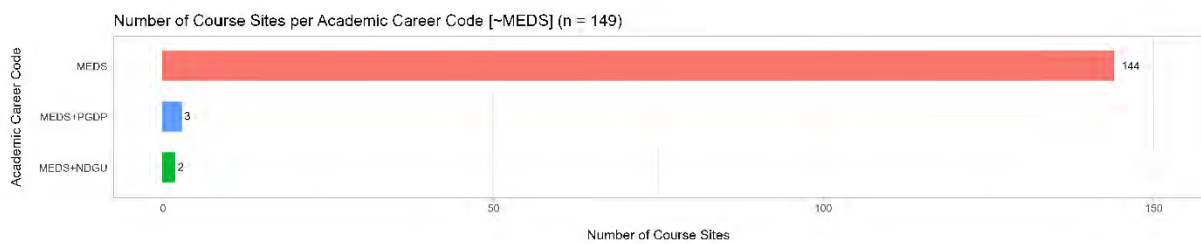


Figure 12: Number of course sites per academic career code filtered on MEDS.

There are three courses in combination with post-graduate degrees (MEDS+PGDP) and 2 with undergraduate non-degree courses (MEDS+NDGU), which are course codes created to manage certain courses for occasional students.

	Title	Faculty	Academic Career	Department	Level	Level #	# Providers
	MBChB Yr4 Obstetrics,2019	FHS	MEDS+NDGU	OBS	4	1	9
	MBChB Yr6 Forensics, 2019	FHS	MEDS+NDGU	MDN+PTY	6	1	5
	MDNP 2019	FHS	MEDS+PGDP	AAE	4	1	1
	PED4049F/S, 2019	FHS	MEDS+PGDP	PED	4	1	2
	MBChB Yr4 ChildHealth2019	FHS	MEDS+PGDP	PED	4	1	1

Table 14: Course sites for MEDS+PGDP or MEDS+NDGU.

Table 14 displays the title, faculty, academic career code, department, year level for the course, number of year levels, and number of provider links for each of the MEDS academic career code combinations. The high number of provider links normally indicates that the course site is used in more of an administrative role instead of teaching and learning. Plotting the number of provider links for all the MEDS academic career codes, it's possible to limit the course sites per number of provider links.

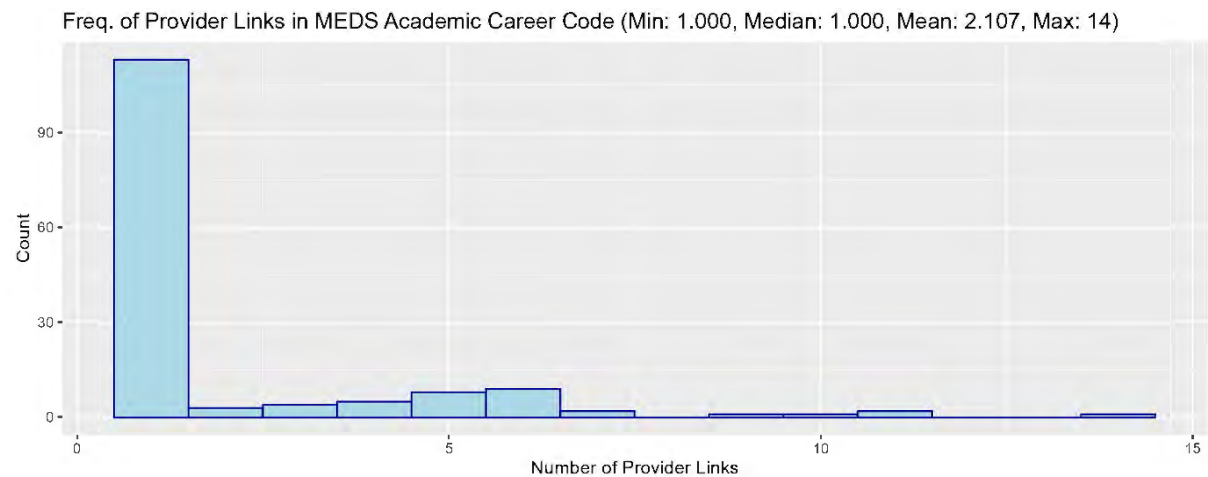


Figure 13: Frequency histogram of provider links in MEDS academic career code.

	Title	Faculty	Academic Career	Department	Level	Level #	# Providers
	CSD All Students,2019	FHS	MEDS	AHS	1+2+3+4	4	14
	MBChB Yr5 Gynae, 2019	FHS	MEDS	OBS	5	1	11
	HS Students Council	FHS	MEDS	NA	NA	0	11
	MBChB Yr5 Ophthalmology	FHS	MEDS	CHM+MDN	5	1	10
	MBChB Yr4 Gen Med 1,2019	FHS	MEDS	MDN	4	1	7
	MBChB Yr6 Fam Med, 2019	FHS	MEDS	PPH	6	1	7
	MBChB Yr5 Ortho Surg 2019	FHS	MEDS	CHM	5	1	6
	MBChB Yr4 Psychiatry,2019	FHS	MEDS	PRY	4	1	6
	MBChB Yr4 Gen Med-2, 2019	FHS	MEDS	MDN	4	1	6
	MBChB5: Care4Children2019	FHS	MEDS	PED	5	1	6

Table 15: MEDS academic career code sites (Top 10) ordered by number of providers.

Reviewing the course sites for the MEDS academic career code strongly suggests that the analysis data set should be filtered on the number of provider links per course site. This will help to exclude administrative sites and sites with multi-focus intent (not teaching and learning).

It is evident (Table 15) that MBCHB (“Bachelor of Medicine and Bachelor of Surgery“ degree) courses make up a majority of these courses and should likely be excluded from the analysis data set. Inspecting the sites confirms that they are used for a collection of courses and materials.

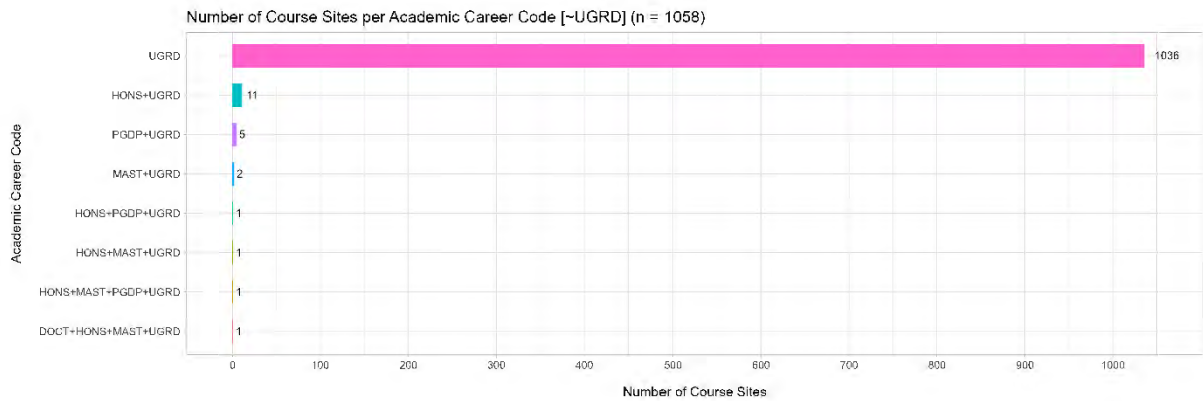


Figure 14: Number of course sites for UGRD academic career code.

Most course sites are just undergraduate sites (1,036). There are additional combinations (7 groups, shown above, 22 sites) with other degree codes. The histogram (Figure 15) based on the number of provider links (below) shows that most sites have a small number of provider links, which fit in with the current constraints.

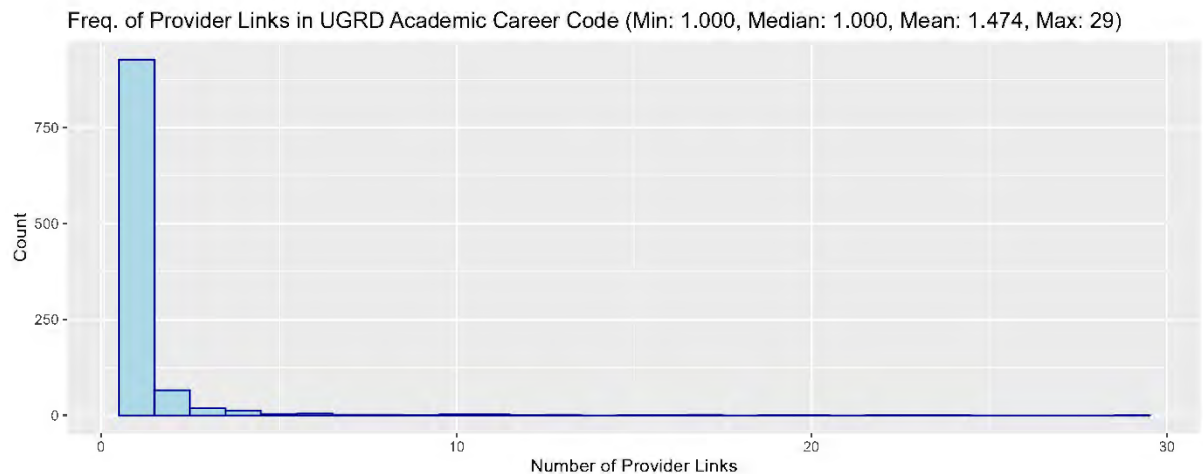


Figure 15: Frequency histogram of provider links in UGRD academic career code.

Title	Faculty	Academic Career	Department	Level	Level #	# Providers
Geomatics Students 2019	EBE	DOCT+HONS+MAST+UGRD	APG	1+2+3+4+5+6	6	24
AcSci Non Academic 2019	COM	HONS+PGDP+UGRD	BUS	1+2+3+4+5	5	17
MFA/BERIndices&Rode	EBE	HONS+MAST+UGRD	CON	2+3+4+5	4	17
3rd Year Dance 2019	HUM	HONS+UGRD	DRM+MUZ	2+3+4	3	11
MSC 2019	HUM	HONS+MAST+PGDP+UGRD	NA	NA	0	10
AcSci 4th Years, 2019	COM	HONS+UGRD	BUS	4	1	8
German Hons 2019	HUM	HONS+UGRD	SLL	3+4	2	6
Performance Making Studio	HUM	HONS+UGRD	TDP	2+3+4	3	6
2019_UD Theory I / ACD	EBE	MAST+UGRD	APG	4+5	2	2
CSC1015F,2019	SCI	MAST+UGRD	CSC	1+5	2	2

Table 16: Combination of UGRD academic career code (Top 10) ordered by number of providers.

In Table 16, there are a few administrative sites, for example, “Geomatic Students 2019” and “MSC 2019”. These course sites have a high number of provider links associated with the course site. This table and Table 15 (MEDS degree courses) suggest using both the number of provider links and the number of year levels to filter out cross-departmental and administrative sites.

Most sites conform to the 2019 selection (year/term) for course sites (Table 18Table 17). The course sites outside the selection criteria contain enrolled students from other years like 2020 and 2018 or, likely in error, 2013, 2015, and 2016.

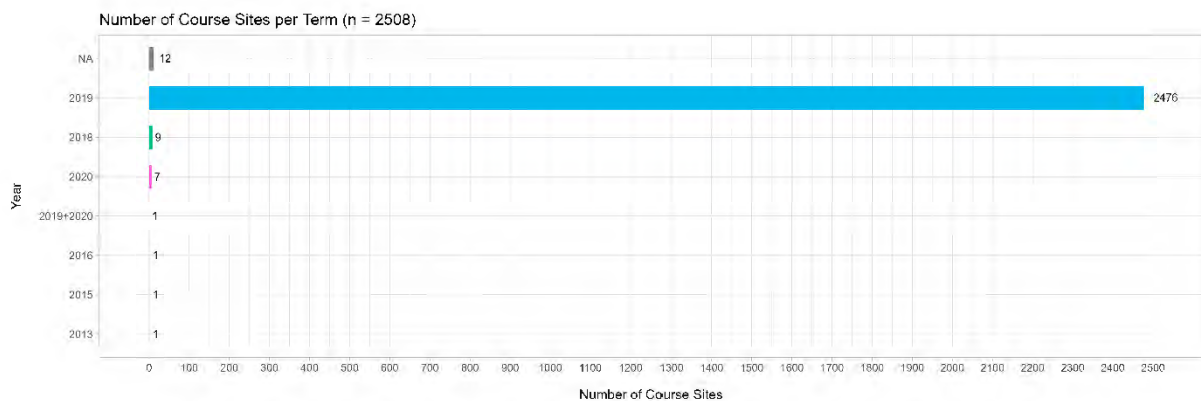


Figure 16: Number of course sites per term.

Title	Faculty	Academic Career	Term	Provider
Turnitin 2019/20: Humanities	NA	NA	NA	Turnitin
Turnitin 2019: Law OLD	NA	NA	NA	Turnitin
Turnitin 2019/20: Commerce	NA	NA	NA	Turnitin
Turnitin 2019: Commerce OLD	NA	NA	NA	Turnitin
FHS Turnitin 2019/20	NA	NA	NA	Turnitin

Table 17: Course sites with NA term

Table 17 has five examples of sites with no year selected; all 12 in the data set are Turnitin sites and are to be excluded from the analysis data set.

Title	Faculty	Academic Career	Term	Provider
GSB MPhil 2016	COM	MAST	2016	GSB5006W,2016
INF5004W,2020 APPLICANTS	COM	MAST	2020	INF_5_004_W_2020
Medicina Forensis, 2020	FHS	MEDS	2020	PTY4008S,2020
MOD 2018/19 RM	COM	BUSN	2018	GSB4616S,2018
EMBA20 2018-2020	NA	NA	2018	NA
PG Dipl 2018/2019	FHS	HONS+PGDP	2018	PPH4051F,2018
BUS4107S, 2018	COM	PGDP	2018	BUS_4_107_S_2019
Electives 2020_MBChB_5&6	NA	NA	2020	NA
2018 Minor Dissertation	COM	BUSN	2018	GSB5117W,2018
CSC1017F,2019	SCI	UGRD	2019+2020	CSC1017F,2019+CSC1017F,2020+CSC1019F,2020
HSE1012S_2018	FHS	MEDS	2018	HSE_1_012_S_2019
BIOSCIENCE HONOURS 2018	NA	NA	2018	NA
PSY5002W-2020	HUM	MAST	2020	PSY_5_002_W_2019
BUS4107, 2018	COM	PGDP	2018	BUS_4_107__2019
PGDip 2018.2 (pre-site)	NA	NA	2018	NA
PHM Registrars	FHS	MAST	2013	PPH7033W,2013+PPH7034W,2013+PPH7035W,2013
GSB MPhil 2015	COM	MAST	2015	GSB5006W,2015
Intro to PG Studies 2020	NA	NA	2020	NA
PSY5003W-2020	HUM	MAST	2020	PSY_5_003_W_2019
MPH 2020 Applications	NA	NA	2020	NA

Table 18: Course sites where the term is not 2019.

Some sites, marked as 2019 course sites, have providers that do not link with the same year (Table 18). The sites are a mixture of application sites for 2020, sites that were not completed in 2019 and continue into 2020, and sites that have been duplicated from previous years. If a site is re-duplicated in 2019, Sakai will set the “term” property value to 2019. The value can be changed later on, but Site Owners often forget. All the course sites will be excluded from the analysis phase as the student activity data will be contaminated by different years.

Before looking at duplication, test, and tutorial sites and also the participant data, the following constraints can be applied:

- I. Course site has a valid Faculty, Department, and Academic Career Code.
- II. Academic Career Code in UGRD, HONS+UGRD, MEDS (focus on undergraduate courses).
- III. Year level between 1 and 3 (exclude sites that span too many years).
- IV. Number Provider links between 1 and 3.
- V. Course site is not for “MBChB”.
- VI. Term is 2019.

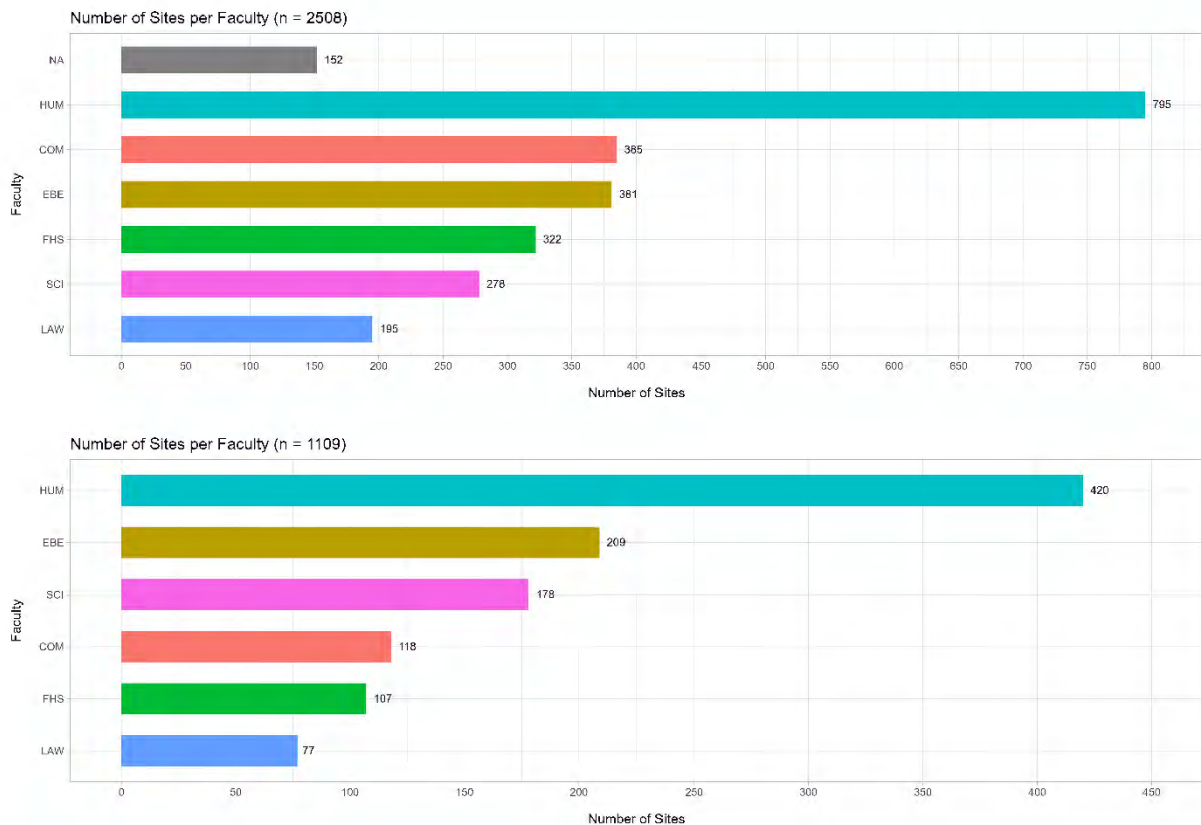


Figure 17: Number of course sites per faculty before and after constraints have been applied.

4.1.6 Exclusion of Duplicate, Test, or Sandbox Sites

After the initial constraints were applied, the course sites were reduced to 1,109. Upon revision of this data set, it became evident that some sites are duplicates of existing sites. This section will identify, evaluate, and, if necessary, exclude these sites.

Lecturers might use these sites to test teaching approaches, assessments, attendance, or content archiving. Typically, the configuration of the provider link for the test sites is different from the live sites, as lecturers would not necessarily want students to be enrolled in them. Table 19 describes the details of sites used for admin queries or managing lab marks and attendance.

Title	Published	gradebookng	overview	resources	roster	siteinfo	qna	search	Tool_Total	User_Total
Admin Queries 2019	TRUE	0	0	0	1	1	1	0	6	1235
Lab Marks Phy1004W 2019	FALSE	1	1	0	1	1	0	1	10	109
Lab Marks PHY1012F,2019	FALSE	1	1	1	1	1	0	1	12	439
Lab Marks PHY1013F,2019	FALSE	1	1	0	1	1	0	1	10	55
Lab Marks PHY1014F,2019	FALSE	1	1	0	1	1	0	1	10	108
Lab Marks PHY1015F,2019	FALSE	1	1	0	1	1	0	1	10	49
Lab Marks PHY1023H,2019	FALSE	1	1	0	1	1	0	1	10	86
Lab Marks PHY1025F,2019	FALSE	1	1	0	1	1	0	1	10	241
Lab Marks PHY1031F,2019	FALSE	1	1	0	1	1	0	1	10	222
Lab Marks PHY1032F,2019	FALSE	1	1	0	1	1	0	1	10	48
Lab PHY2004W,2019	FALSE	1	1	1	1	1	0	1	12	64
STA1000F Admin Queries 19	TRUE	0	1	0	0	1	1	0	6	285

Table 19: Course site with low total tool count, exception sites

The site “Admin Queries 2019” only has one student visible tool assigned, and the similarly named site “STA1000F Admin Queries 19” follows the same structure. These course sites were created in February and July of 2019 (first and second semesters) for the Department of Statistical Sciences 1st year courses. The sites have a limited tool set, primarily “Question and Answers” (qna) and a reasonably high student-to-lecturer ratio (1225:1 and 227:2, respectively). The “Question and Answers” (qna) tool was used to answer questions students might have about assignments, lectures, quizzes, tests, tutorials, workshops, and other topics. The assessment type tool was used as a Frequently Asked Questions (FAQ) communication tool.

The Department of Physics course sites with the title structure “Lab [Marks] [Provider]” include a couple of tools, namely Overview (a welcome page), Gradebook, Roster, and Search. The concluded purpose of this site is to display the list of participants (roster) and allow the lecturer to communicate the students' respective grades for each of their Lab sessions. Although “Lab PHY2004W,2019” and “Lab Marks PHY1012F,2019” have the resources tool linked, there are no resources linked (no files uploaded).

These sites were all excluded from the data set as they are used in an administrative capacity to communicate grades and attendance at laboratory sessions.

Identifying and excluding sites with “sandbox”, “test”, or “duplicate” in the title is a straightforward process, but it's not always possible to find every site with these keywords. The sites containing these keywords are shown in Table 20.

Title	Published	Tool Total	User Total	Created On
ACC2112W,2019(sandbox)	TRUE	18	2	2019-01-24 11:15:24
AGE2012F,2019sandbox	TRUE	8	3	2019-01-24 11:05:54
AGE3013H,2019(sandbox)	TRUE	19	4	2019-01-24 11:02:49
AGE3013H,2019(sandbox)	TRUE	21	2	2019-01-24 11:03:04
BUS1036F 2019 (test site)	TRUE	38	7	2019-05-16 21:56:16
FTX3045S- TEST 1 - 2019	FALSE	5	432	2019-08-23 15:21:46
FTX3045S TEST 1	FALSE	5	19	2019-08-23 15:17:20
INF1102S_Makeup Test	FALSE	4	151	2019-09-23 12:59:13
INF3014F_Test 2, 2019	FALSE	6	83	2019-05-03 11:57:14
Tests STA1000S,2019	TRUE	8	1236	2019-08-12 11:08:18

Table 20: Sites filtered by “test” or “sandbox” in the title.

The site “INF1102S_Makeup Test” is set up for Informatics students to do a make-up test. The site “Tests STA1000S,2019” is used to administer a Quiz to students; both these sites are valid and should not be excluded. “FTX3045S- TEST 1 – 2019” and “FTX3045S TEST 1,” on the other hand, do have corresponding teaching courses.

Title	Published	Tool Total	User Total	Created On
FTX3045S- TEST 1 - 2019	FALSE	5	432	2019-08-23 15:21:46
FTX3045S-2019	TRUE	17	466	2019-06-10 09:26:40
FTX3045S TEST 1	FALSE	5	19	2019-08-23 15:17:20

Table 21: Test sites for FTX3045S.

Similarly, “BUS1036F,2019” has a test site “BUS1036F 2019 (test site)”, which will be excluded.

These sites are sometimes not published (made available to students). There are appropriate reasons for a lecturer to unpublish their sites, which alone is not a good indicator of test sites. Further title analysis is required to find duplicate sites (if obligingly indicated by the lecturer's change of site title), test, or sandbox sites.

Title	Provider	Published	Created On	Tool Total	User Total
AGE3013H,2019(sandbox)	NA	TRUE	2019-01-24 11:02:49	19	4
AGE3013H,2019(sandbox)	NA	TRUE	2019-01-24 11:03:04	21	2
AHS1003F,2019	AHS1003F,2019	TRUE	2018-10-18 10:42:14	13	57
AHS1003F,2019	NA	TRUE	2019-01-18 15:31:54	10	3
BUS2022S,2019	BUS2022S,2019	TRUE	2019-06-13 13:30:40	14	179
BUS2022S,2019	NA	FALSE	2019-01-11 09:44:19	3	2
CML4506F,2019	NA	FALSE	2019-02-05 09:52:47	9	1
CML4506F,2019	CML4506F,2019	TRUE	2019-02-04 10:18:06	9	45
CML4607Z,2019	NA	TRUE	2019-02-05 10:09:30	9	1
CML4607Z,2019	CML4607Z,2019	TRUE	2019-02-04 10:36:42	10	183
CON1019S,2019	NA	FALSE	2019-06-19 14:50:33	1	1
CON1019S,2019	CON1019S,2019	TRUE	2019-06-19 15:39:47	13	45
EEE3093S,2019	EEE3093S,2019	FALSE	2019-02-17 11:41:50	8	52
EEE3093S,2019	EEE3093S,2019	TRUE	2019-07-08 10:37:14	12	61
FIN1009S,2019	FIN1009S,2019	TRUE	2019-07-11 11:56:36	10	115
FIN1009S,2019	FIN1009S,2019	FALSE	2019-01-21 11:43:56	9	106
FIN2028S,2019	FIN2028S,2019	TRUE	2019-07-11 12:00:29	11	81
FIN2028S,2019	FIN2028S,2019	FALSE	2019-01-21 12:06:27	8	76
FIN3027S,2019	NA	FALSE	2019-01-21 14:20:19	8	6
FIN3027S,2019	FIN3027S,2019	TRUE	2019-07-11 12:05:28	10	42
FIN3029S,2019	FIN3029S,2019	TRUE	2019-07-11 12:06:21	10	28
FIN3029S,2019	FIN3029S,2019	FALSE	2019-01-21 14:16:31	8	26
HSE1012S,2019	HSE1012S,2019	TRUE	2019-07-16 13:40:15	9	8
HSE1012S,2019	NA	TRUE	2019-11-11 10:45:51	8	7
PRY2002W,2019	PRY2002W,2019	TRUE	2019-01-17 13:33:37	14	47
PRY2002W,2019	PRY2002W,2019	TRUE	2019-03-08 14:35:30	9	42
RDL1006W,2019	RDL1006W,2019	TRUE	2019-01-30 11:31:30	10	25
RDL1006W,2019	RDL1006W,2019	TRUE	2019-02-04 11:25:10	9	20
SLL1132S,2019	SLL1132S,2019	TRUE	2019-07-09 09:40:59	13	21
SLL1132S,2019	SLL1132S,2019	TRUE	2019-06-28 16:01:07	12	23
SWK1013F,2019	NA	FALSE	2019-02-21 14:56:01	18	1
SWK1013F,2019	SWK1013F,2019	TRUE	2019-02-04 16:44:14	14	331
SWK2013S,2019	SWK2013S,2019	TRUE	2019-02-22 12:46:50	8	162
SWK2013S,2019	NA	FALSE	2019-01-24 11:01:32	9	3
SWK4015F,2019	SWK4015F,2019	TRUE	2018-11-16 14:46:37	11	56
SWK4015F,2019	SWK4015F,2019	TRUE	2019-01-18 13:38:22	4	54
SWK4031S,2019	SWK4031S,2019	TRUE	2019-06-01 20:09:14	9	53
SWK4031S,2019	SWK4031S,2019	TRUE	2019-01-18 12:19:41	5	51

Table 22: Course sites with the same title.

There are 38 sites with an exact match on the title; some sites can be excluded because they are sandbox sites, have been unpublished, or do not have a provider link. Five sites warrant further investigation (Table 23).

Title	Provider	Published	Created On	Tool Total	User Total
PRY2002W,2019	PRY2002W,2019	TRUE	2019-01-17 13:33:37	14	47
PRY2002W,2019	PRY2002W,2019	TRUE	2019-03-08 14:35:30	9	42
RDL1006W,2019	RDL1006W,2019	TRUE	2019-01-30 11:31:30	10	25
RDL1006W,2019	RDL1006W,2019	TRUE	2019-02-04 11:25:10	9	20
SLL1132S,2019	SLL1132S,2019	TRUE	2019-07-09 09:40:59	13	21
SLL1132S,2019	SLL1132S,2019	TRUE	2019-06-28 16:01:07	12	23
SWK4015F,2019	SWK4015F,2019	TRUE	2018-11-16 14:46:37	11	56
SWK4015F,2019	SWK4015F,2019	TRUE	2019-01-18 13:38:22	4	54
SWK4031S,2019	SWK4031S,2019	TRUE	2019-06-01 20:09:14	9	53
SWK4031S,2019	SWK4031S,2019	TRUE	2019-01-18 12:19:41	5	51

Table 23: Duplicate titles that are published with a valid provider.

Looking at the students enrolled in each site and the events generated (more detail later), Table 24 shows which sites to exclude from the data set.

COURSE SITE	# STUDENTS	STUDENTS	# EVENTS	% DIFF	
PRY2002W,2019	41	The same students in both sites	A	9,498	125.107%
			B	2,188	
RDL1006W,2019	17	The same students in both sites	A	675	24.813%
			B	526	
SLL1132S,2019	17	The same students in both sites	A	2,697	198.262%
			B	618	
SWK4015F,2019	52	The same students in both sites	A	14,090	194.062%
			B	935	
SWK4031S,2019	49	The same students in both sites	A	10,676	195.129%
			B	866	

Table 24: Comparison of duplicate sites showing the number of students and events.

All the sites have provider links, so the student enrolment for each is identical. The events are the sum of all events for students in those sites (student interaction with tools in the site), which in all the B sites are lower than the A sites. Calculating the percentage difference between the number of events for sites A and B ($|A-B|/[A+B/2] \times 100$), the sites with a significant difference will be excluded, which in this case means that both “RDL1006W,2019” sites will remain in the courses data set.

To find duplicate sites, the course titles can be analysed (as shown above), allowing for situations where the titles might differ slightly. To achieve this, several string comparison methods can be used. These methods include Levenshtein, Damerau-Levenshtein, Jaro-Winkler, Cosine similarity, and REGEXP. While each method has its strengths, REGEXP was found to be the most accurate in detecting duplicate course sites.

In reviewing *all* the sites in the data set and evaluating them in a similar manner as the sites in Table 24 (i.e., by looking at tools, participants, and event data), 66 course sites were identified as duplicates. They were excluded from the final data set.

4.1.7 Understanding Site Participants

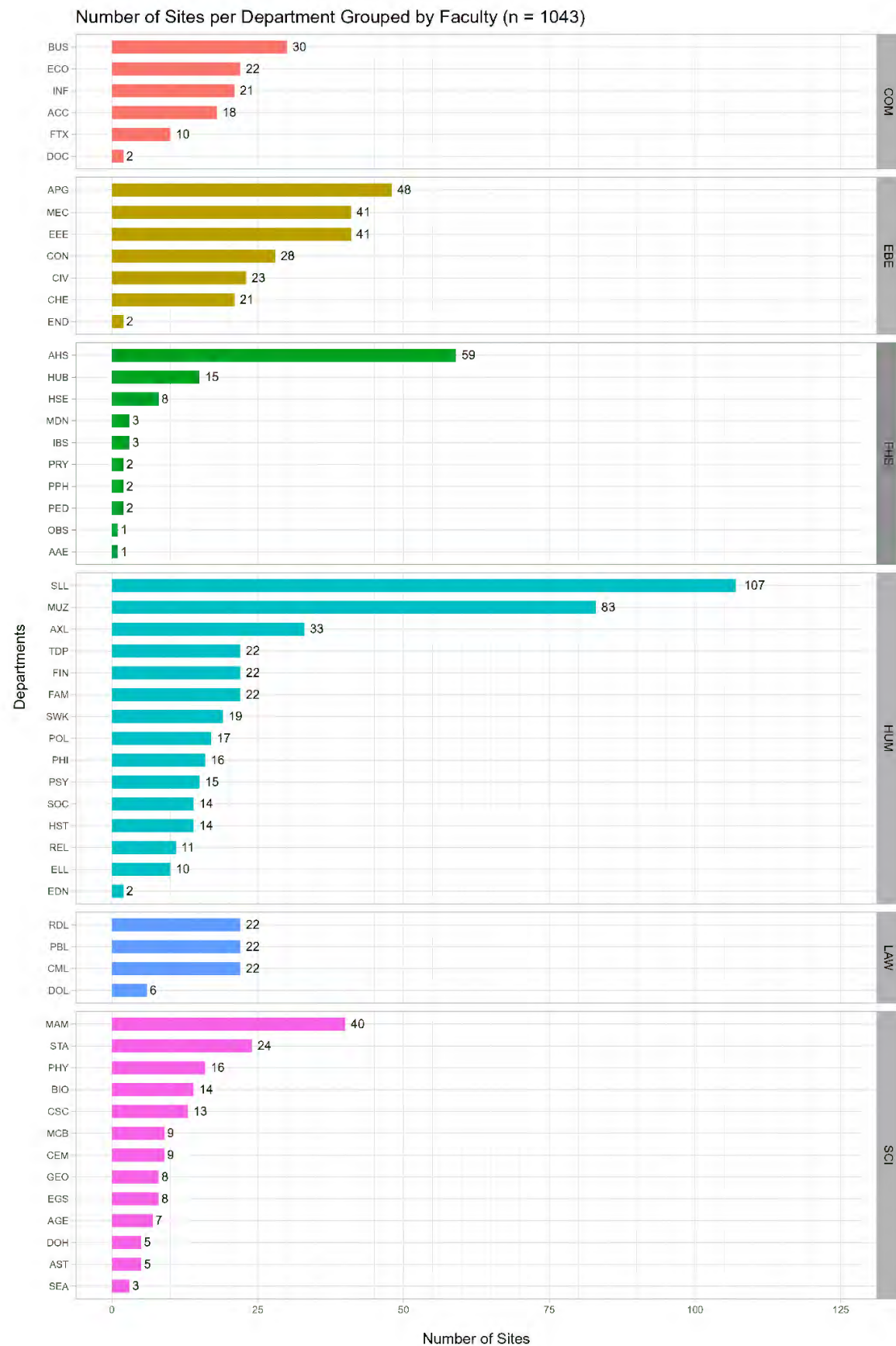


Figure 18: Number of course sites per department grouped by faculty.

Figure 18 reveals that out of the 1,043 course sites, the Faculty of Humanities (HUM: 407) has the most course sites, followed by the Faculty of Engineering & Built Environment (EBE: 204), then the Faculty of Science (SCI: 161), Faculty of Commerce (COM: 103), second to last Faculty of Health Sciences (FHS: 96) and then lastly Faculty of Law (LAW: 72).

In the next section of this document the enrolment (or participants as it is called in Sakai) will be examined as this might also impact the tool choices for the course sites.

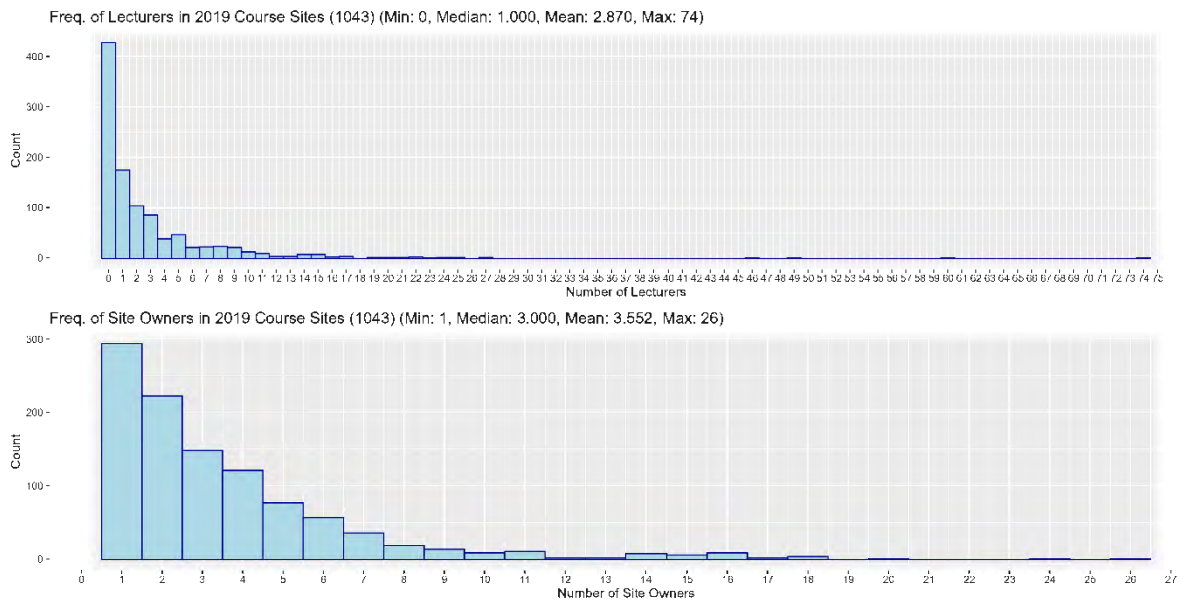


Figure 19: Histogram of the number of Lecturers and Site Owners in 2019 course sites

In Sakai, when a user creates a new course site, that user will automatically be enrolled as a “Site Owner” in the site; this means there will *always* be such a role for the course site. The “Site Owner” role is practically identical to the “Lecturer” role, as it just differs by name; most users don’t change their roles. The user might not be the eventual lecturer for the course site and might be part of the administration staff of a department used to create course sites. Looking at the histogram of “Lecturer” roles in course sites, it is clear that there are quite a number of sites (429) that do not have that role assigned, but there are no sites without a “Site Owner”. Table 25 shows five examples of such sites.

Title	Lecturer	Site Owner	Support Staff	Tutor	Student	Other
3MP, 2019	0	1	0	0	28	0
ACC2023F 2019	0	9	3	7	235	2
AGE1004S,2019	0	3	0	0	22	0
AGE3012S,2019	0	1	0	0	14	0
AHS1003F,2019	0	4	2	0	51	0

Table 25: Course site where Lecturer participants are 0.

In the further mapping of participants, the “Lecturer” and “Site Owner” roles will be combined as they have the same permissions and are interchangeable in course sites.

On the other end of the histogram, there are sites with many “Lecturer” or “Site Owners”.

Title	Lecturer	Site Owner	Support Staff	Tutor	Student	Other
ECO3025S,2019	0	26	0	0	79	2
ECO3021S,2019	0	24	0	5	271	7
ACC4000H,2019	8	20	2	0	313	5
ECO2003F,2019	0	18	0	23	813	2
BUS1003H,2019	0	18	0	14	187	1
BUS4027W & BUS4088S,2019	0	18	0	4	79	0
ECO1010F,2019	1	18	1	42	1125	16
ECO3020F,2019	0	17	0	15	373	1
ELL3001S,2019	0	17	2	0	202	1
ELL1016S,2019	1	16	0	16	391	1

Table 26: Course site participants ordered by number of Site Owner descending (Top 10)

The sites shown in Table 26 emphasise that combining the “Site Owner” and “Lecturer” is a valid approach. Reviewing the top course sites by “Lecturer” (table below) firstly highlights the chance that a site like “AHS4065W,2019”, which has 35 participants (“Lecturer”: 27, “Site Owner”: 3, “Support Staff”: 5) has no students (“Student”: 0). It also shows that there are sites with a large number of students versus lecturers, like “Discovery Commerce 2019” with 25 “Lecturer”, 16 “Site Owner” and 1,107 “Student” roles.

Title	Lecturer	Site Owner	Support Staff	Tutor	Student	Other
CHE4045Z,2019	74	3	6	0	90	0
IBS1007S, 2019	60	7	3	25	231	1
Molecular Medicine 2019	49	2	1	0	17	0
EEE4022S,2019	46	2	1	0	117	0
AHS4065W,2019	27	3	5	0	0	0
CHE2005W,2019	27	10	3	28	109	0
HUB1006F, 2019	25	3	5	18	227	1
Discover Commerce 2019	25	16	0	46	1107	106
EEE4022F,2019	24	2	2	0	18	0
AHS4121W,2019	24	6	4	0	70	0

Table 27: Course site participants ordered by number of Lecturers descending (Top 10)

The next set of participants to analyse before the “Student” role are “Support Staff” and “Tutors”, the roles that support the teaching and learning on the site. These roles have more access to the various tools (can mark assignments, upload content, and create announcements) but cannot add or remove tools and participants.

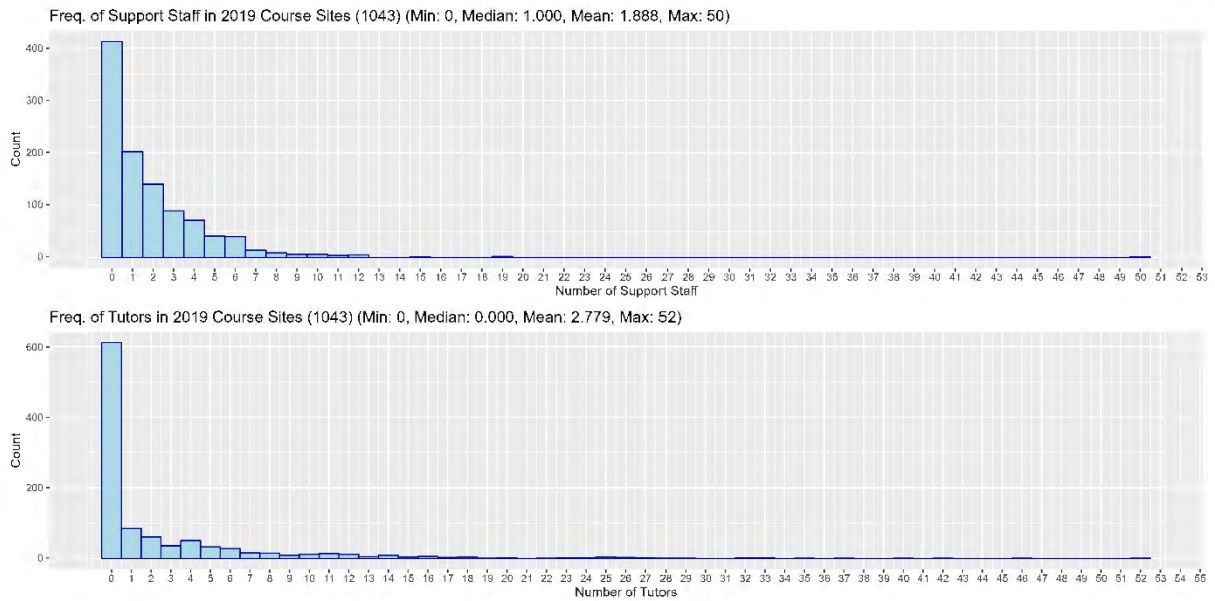


Figure 20: Histogram of the number of Support Staff and Tutors in 2019 course sites

The histograms describe the number of “Support Staff” and “Tutor” roles in the course sites. More than a third of the sites do not have any of these roles assigned to them.

As these roles are close to identical in the LMS, their combined totals can be viewed to get a better sense of how the support roles are distributed in the course sites.

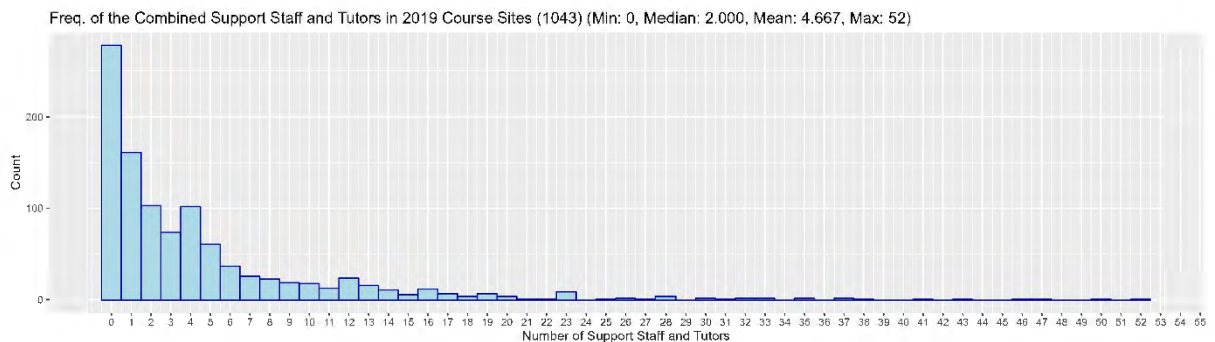


Figure 21: Histogram of the number of Support Staff and Tutors (combined) in 2019 course sites

The histogram (skewed right) shows the expected result that most course sites do not include support staff (280), which means that they rely more on lecturers’ interactions. Looking at sites with a high number of support staff (Table 28), it is easy to see that the participants were added due to the high number of students in the course site.

Title	Lecturer	Site Owner	Student	Support Staff	Tutor	Combined
ECO1011S 2019	1	15	1423	0	52	52
CSC1016S,2019	15	2	590	50	0	50
PSY1004F - 2019	6	2	1002	7	40	47
Discover Commerce 2019	25	16	1107	0	46	46
ECO1010F,2019	1	18	1125	1	42	43
PPH1001F 2019	16	4	366	4	37	41
STA1000S,2019	4	1	1227	9	29	38
MAM1000W,2019	6	1	580	2	35	37
PSY1005S-2019	6	6	726	5	32	37
CSC2001F,2019	14	1	265	10	25	35

Table 28: Course site participants ordered by number of Support Staff and Tutors (Top 10)

There are other possible participants (roles) that can be added to a course site. These roles are “Reviewers”, “Observers”, and “Librarian”.

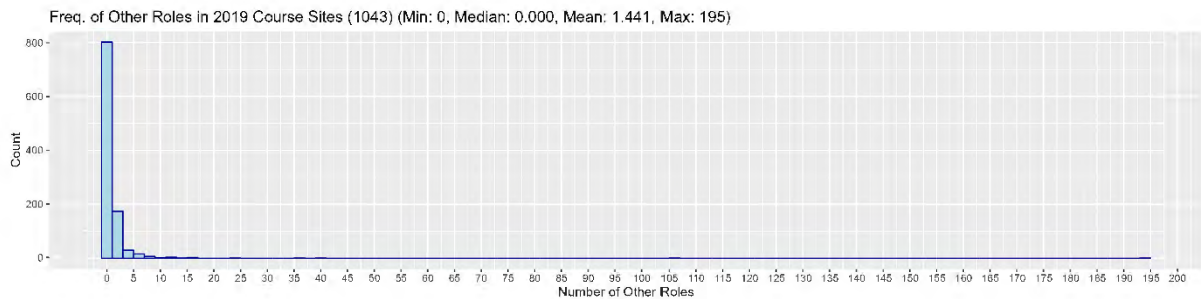


Figure 22: Histogram of the Other Roles (Reviewers, Observers, and Librarian) in 2019 Course Sites

The histogram (Figure 22) shows that there are a few courses that include these roles, but mostly in small numbers; the outliers (shown in the table below) indicate that there are a large number of “Observer” roles in these sites. The “Observer” role is a read-only access role that can access content but has limited interaction with other site participants. In a sense, these roles are all excluded from this analysis and are shown here for completeness.

Title	Lecturer	Site Owner	Student	Observer	Reviewer	Librarian	Other
ACC2012W 2019	19	12	311	195	0	0	195
Discover Commerce 2019	25	16	1107	104	2	0	106
FTX2024S-2019	1	4	613	40	0	0	40
CEM1000W 2019	7	2	381	37	0	0	37
CEM1011F,2019	1	3	212	24	0	0	24
CSC2002S,2019	15	1	250	17	0	0	17
CHE4067F,2019	5	1	32	16	0	0	16
ECO1010F,2019	1	18	1125	15	0	1	16
CSC2001F,2019	14	1	265	14	0	0	14
HUB1019F,2019	9	1	101	13	0	0	13

Table 29: Course site participants ordered by number of “Other” (Top 10)

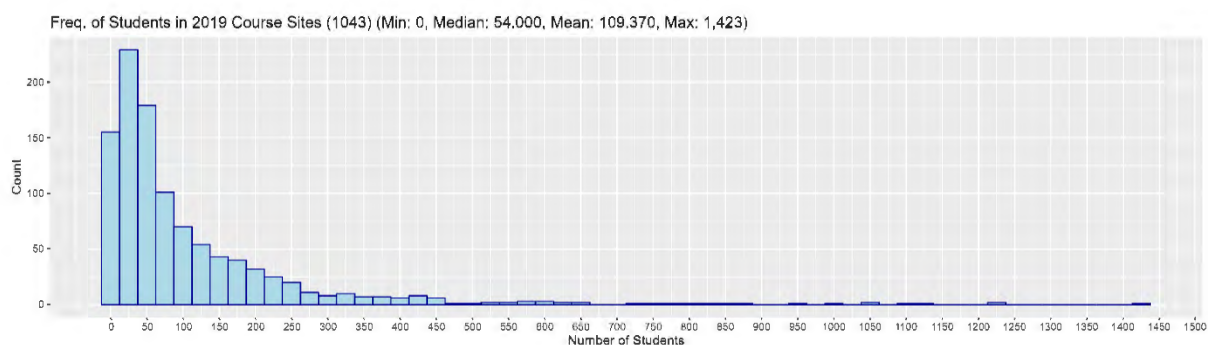


Figure 23: Histogram of the Number of Students in 2019 Course Sites

The right-skewed histogram of the “Student” role in course sites indicates that most sites contain a low number of students, but there are sites that contain significantly more students than average (Max: 1,423). Table 30 contains the top 10 sites by student enrolment.

Title	Lecturer	Site Owner	Support Staff	Tutor	Student	Other
ECO1011S 2019	1	15	0	52	1423	4
STA1000S,2019	4	1	9	29	1227	7
Tests STA1000S,2019	3	1	6	0	1226	0
ECO1010F,2019	1	18	1	42	1125	16
Discover Commerce 2019	25	16	0	46	1107	106
CSC1015/17 Challenge,2019	14	2	5	0	1057	0
ACC1006F 2019	4	9	4	22	1040	5
PSY1004F - 2019	6	2	7	40	1002	2
BUS1036F,2019	0	2	5	14	942	3
SOC1001F,2019	5	4	5	12	879	0

Table 30: Course site participants ordered by Number of Students (Top 10)

There are a couple of interesting course sites shown in the table; for example, “Tests STA1000S, 2019” is a site with a student-to-lecturer ratio of 306.50 (1226:4) (on the higher end of the scale) and a student-to-support (tutor) ratio of 204.33 (1226:6) (also on the higher end of the scale). This site contains 8 tools, 4 visible to students (Announcements, Test & Quizzes, Test Group Signups). This site was created and used primarily for the exam of this course:

EXAM INSTRUCTIONS:

DO NOT OPEN TESTS|&QUIZZES UNTIL THE CHIEF INVIGILATOR TELLS YOU THAT YOU MAY BEGIN THE EXAM!

Figure 24: Course site overview of “Tests STA1000S,2019” - indicating primary usage.

The site with the highest number of students (1,423) is “ECO1011S 2019”, with a student-to-lecturer ratio of 88.938 (1423:16) and a student-to-support ratio of 27.365 (1423:52). The site is used to teach the introduction to macroeconomics to 1st year students.

Another interesting site with a high number of students (1,107) is the Faculty of Commerce site “Discover Commerce 2019”, linked to the course codes DOC1103H and DOC1003H (Providers). This course site is constructed with several lesson tools, resources, and communication tools to support

the student in creating “an excellent foundation for success, both at UCT, and in your future lives and careers”.

≡ OVERVIEW

Welcome to all Commerce First Year students

DISCOVER COMMERCE

This DISCOVER COMMERCE Vula site has been created just for you, and the various elements of this site are designed to do several things to assist you create an excellent foundation for success, both at UCT, and in your future lives and careers.

There are two key aspects to this site:

1. [The Commerce Toolkit](#)
2. [The Commerce Case Study](#)

The **Commerce Toolkit** and **Commerce Case Study** contain the content for the courses DOC1003H and DOC1103H. If you are registered for a mainstream degree in Commerce, you will be registered for DOC1003H; if you are registered in EDU, you will be registered for DOC1103H. Please note that the course counts 5 credits towards your degree, and it is compulsory to pass this course. This means that all Commerce students will be required to pass this course in order to graduate.

Figure 25: Course site overview of "Discover Commerce 2019" - indicating primary usages.

The Department of Computer Science created a course site for all 1st year students enrolled in CSC1015F and CSC1017F (Providers); it has 1,057 students enrolled and contains custom learning tools to provide an environment for students to expand their programming skills in Python.

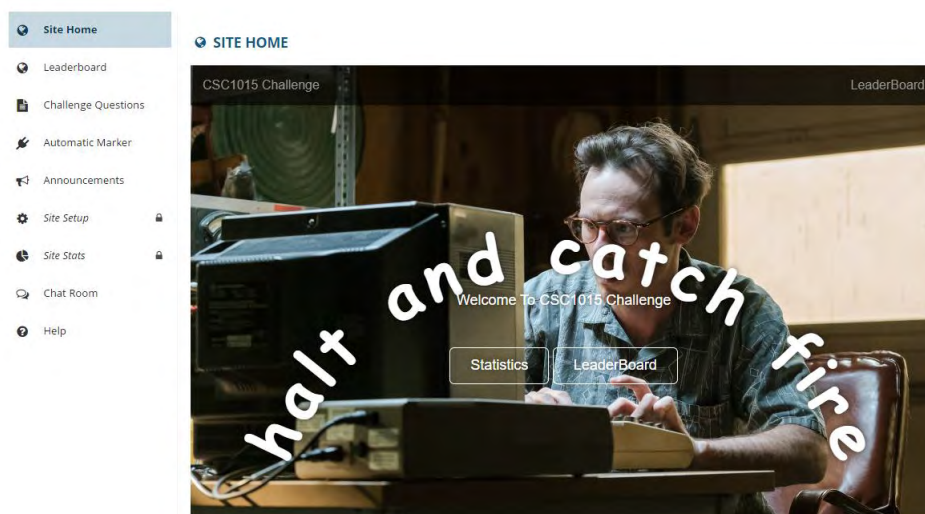


Figure 26: Course site overview of "CSC1015/17 Challenge,2019"

Focusing on the sites with a low number of students, the histogram (Figure 27) has more of a multimodal distribution and indicates that there are some sites with no students at all.

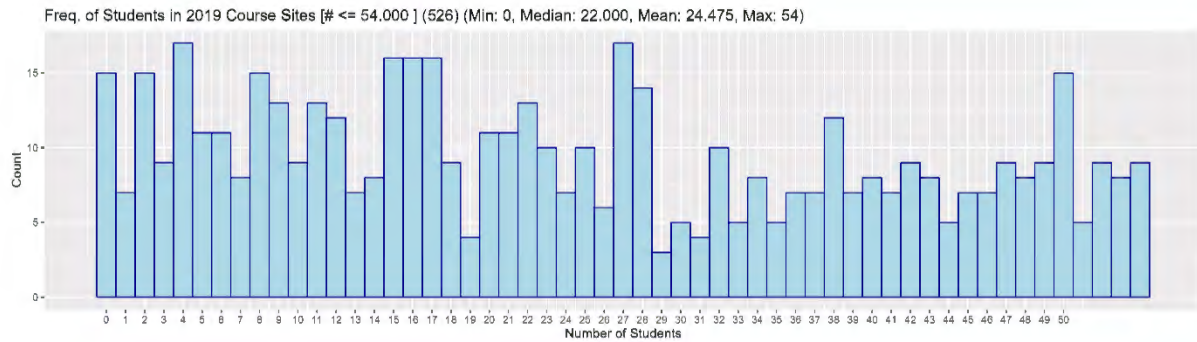


Figure 27: Histogram of the number of students in 2019 course sites with 50 or fewer students

No. Sites	No. Students
15	0
7	1
15	2
9	3
17	4
11	5
11	6
8	7
15	8
13	9
9	10

Table 31: Number of Sites per Number of Students (Bottom 10)

Table 31 is an extract of the histogram data that shows the number of sites that contain the lowest number of students (counting the number of sites with 10 or fewer students).

This highlights an issue which will impact the final assessment, in which the student interactions with a course site will be evaluated. The sites with a low number of students (0, 1, and/or 2) should probably be excluded from the analysis data set as the results obtained from them would likely be inaccurate when applied to the complete student body.

Table 32 shows the enrolment details for course sites in which there are no students enrolled.

Title	Lecturer	Site Owner	Combined	Student	Support Staff	Tutor
AHS4065W,2019	27	3	30	0	5	0
AXL3200F,2019	2	3	5	0	0	0
MEC4104F,2019	0	2	2	0	1	0
MUZ4371H,2019	0	2	2	0	0	0
RDL4606F,2019	0	3	3	0	0	0
PRY4000W,2019	0	6	6	0	0	0
Hebrew IA,2019	0	5	5	0	0	0
EEE1000X,2019	0	1	1	0	1	0
City Ancient World,2019	0	2	2	0	0	0
Greek II & IIIB,2019	0	3	3	0	0	0
SLL1046S,2019	0	1	1	0	0	0
PBL4601S,2019 (ConLit)	0	2	2	0	0	0
BUS2035S,2019	0	1	1	0	2	0
PED5004W,2019	0	2	2	0	0	0
Jazz Improv 4 SSA, 2019	0	1	1	0	0	0

Table 32: Course site with 0 students (15)

Title	Lecturer	Site Owner	Combined	Student	Support Staff	Tutor
MUZ2821H, 2019	1	1	2	1	2	0
PSY2014S term 4 module	1	3	4	1	0	0
Piano D1, 2019	0	3	3	1	0	0
PSY2014S term 3 module	0	4	4	1	0	0
Greek II & IIIA, 2019	0	6	6	1	0	0
African Aural 2 MUZ2380	0	1	1	1	0	0
MUZ2278H, 2019	0	1	1	1	1	0
MUZ1278H, 2019	1	1	2	2	0	0
HSE1013F, 2019	1	2	3	2	0	0
MUZ4278W, 2019	0	1	1	2	1	0
MUZ1382H, 2019	0	4	4	2	0	0
DANCE TEACHING METHODS II	0	1	1	2	0	0
MUZ2373H, 2019	0	3	3	2	0	0
MUZ3278H, 2019	0	1	1	2	1	0
Italian IIIB, 2019	0	5	5	2	0	0
Hebrew IB SLL1083S, 2019	0	3	3	2	0	0
RDL4609H, 2019	0	1	1	2	0	0
Italian IIIA, 2019	0	6	6	2	0	0
MUZ3372H, 2019	0	3	3	2	0	0
musicology 2019	0	1	1	2	0	0
Latin II & IIIB, 2019	0	7	7	2	0	0
Jazz Singing B4, 2019	0	1	1	2	0	0

Table 33: Course site with 1 or 2 Students (7/15)

As stated before, the number of students in a site is important because their activity on the site will be used to determine the efficacy of the tools used. With this in mind, the constraint will be updated to include all course sites with at least 4 students enrolled.

For the following set of graphs, this restriction is applied, and the “Lecturer” and “Support Staff” role was combined into “Lecturer”; the same was done for “Support Staff” (which is the combination of “Support Staff” and “Tutor”). Figure 28 is the updated histogram with the sites removed with less than 4 students.

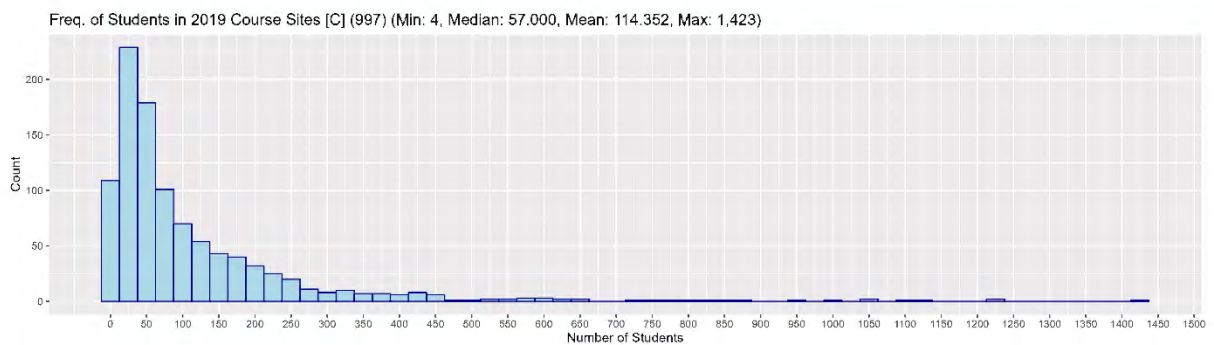


Figure 28: Histogram of the Number of Students in 2019 Course Sites (with more than 3 students).



Figure 29: Histogram of the Number of Students in 2019 Course Sites

The histogram (Figure 29) shows the ratio of students to lecturers for the 997 course sites. This ratio is generally quite low (mostly below 50 students to lecturer); there are some outliers, which, when examined, might add to the existing constraints. The following table (Table 34) shows the top 10 course sites with the highest student-to-lecturer ratio.

Title	Department	Faculty	Lecturer	Student	Ratio	Support Staff	# Tools
CIV2020X,2019	CIV	EBE	1	554	554.0	0	8
BUS1036F,2019	BUS	COM	2	942	471.0	19	33
Tests STA1000S,2019	STA	SCI	4	1226	306.5	6	8
STA1000S,2019	STA	SCI	5	1227	245.4	38	26
MEC1000X,2019	MEC	EBE	1	240	240.0	2	8
MDN3001H,2019	MDN	FHS	1	231	231.0	2	8
MAM2084S,2019	MAM	SCI	2	419	209.5	5	11
MEC1003F,2019	MEC	EBE	1	176	176.0	14	13
MAM1010F,2019	MAM	SCI	5	851	170.2	25	14
MAM2083F, 2019	MAM	SCI	2	322	161.0	4	18

Table 34: Course site order by Student to Lecturer ratio (Top 10)

The highest ratio is found on the site for “CIV2020X,2019”. This site is used primarily to communicate with students and deliver resources for their studies, so although it has a high student-to-lecturer ratio and learning content is scarce, it still qualifies as a teaching course site. In the top 3 is “Tests STA1000S,2019” (which was discussed previously) and “CIV2020X,2019”. Both have a high student-to-lecturer ratio and the tools selected are centred around signing up for groups, and announcements for the “Test & Quizzes” tool. The other tools are lecturer visible tools, which help with administration like gradebook and site setup. In contrast, the site “STA1000S,2019” is a well-structured teaching site with tools to communicate, assess and deliver content to students. “BUS1036F,2019” is similarly structured.

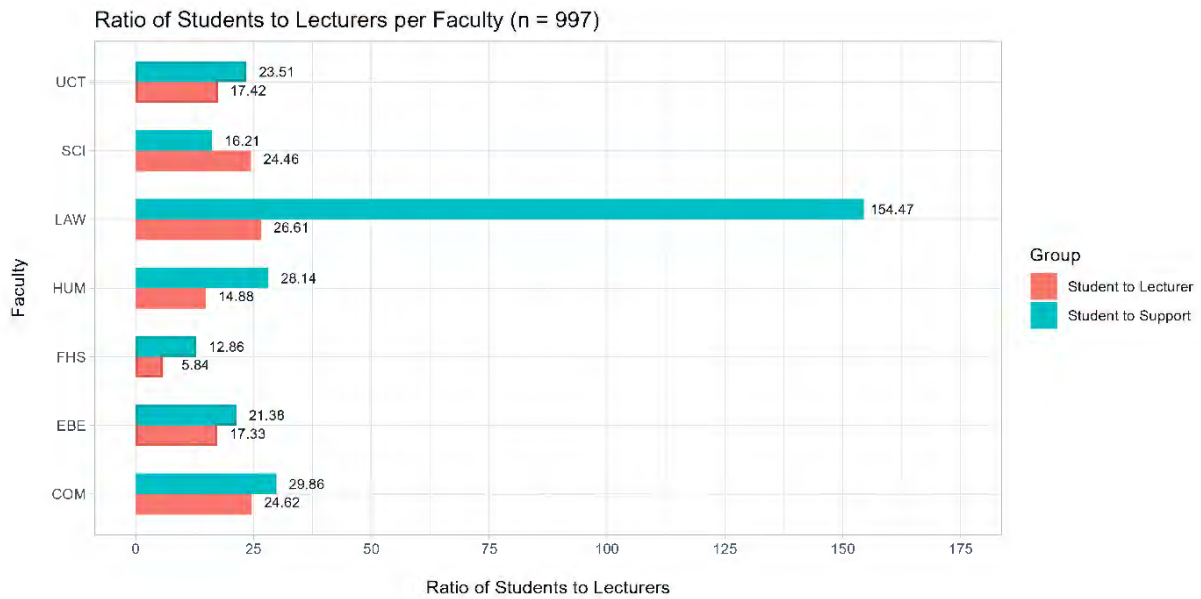


Figure 30: Ratio of Student to Lecturer and Students to Support by Faculty.

Figure 30 shows the combined “Student to Lecturer” ratio and “Students to Support” for each faculty for the chosen undergraduate sites. To note is that the number of “Lecturer” users used in the ratio calculation is the combined roles of “Lecturer” and “Site Owner”, and the “Support” is the combined value of “Support Staff” and “Tutor”. The actual values for each faculty might differ somewhat from those reported in the extract from the UCT Teaching and Learning Report (Various, 2019). This shows the full-time equivalent enrolments (FTE) for academic staff grouped by faculty. The “Student to Lecturer” (student to academic staff) differs from the official figures because not all academic staff teach in the LMS; the site owners (which might be administrative staff) are combined with the “Lecturer” value (in the data set), and the site selection is limited to undergraduate courses.

Figure 8: Weighted FTE enrolments per academic staff member: 2015 and 2019

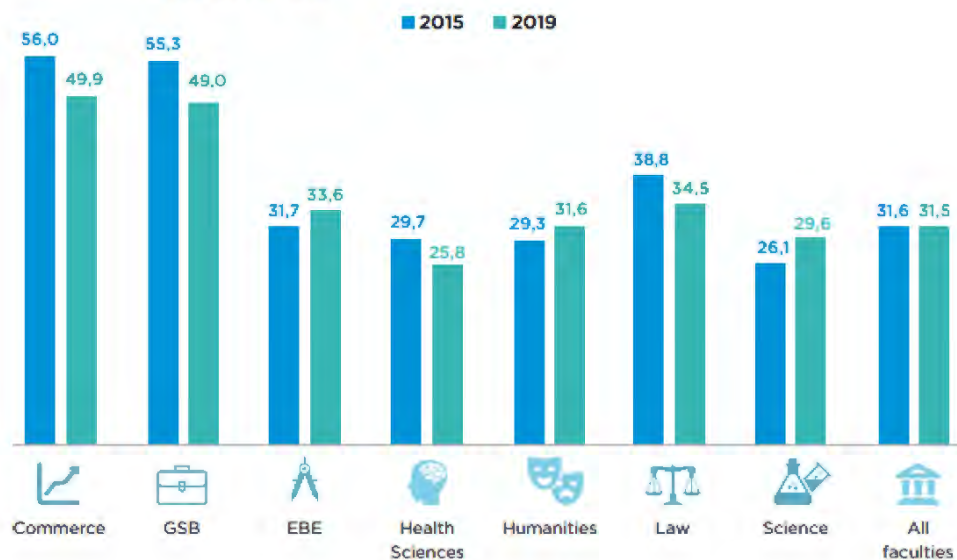


Figure 31: Full-time equivalent (FTE) enrolments per academic staff member across the institution.

Plotting the number of students to the number of lecturers (Figure 32 - grouped by faculty) shows that the ratio (x/y) of student to lecturers are generally quite low (clustered around the low X and Y values). The outliers are for courses in the Faculty of Commerce (COM) where there are a high number of students and a low number of lecturers (i.e., “ECO1011S 2019”, “ECO1010F,2019”, “Discover Commerce 2019”). The Faculty of Science (SCI) courses are ones that were identified already (“Tests STA1000S,2019”, “STA1000S,2019”, “CSC1015/17 Challenge,2019”). There aren’t any courses that fall into the quadrant of many students with many lecturers (top right). Some courses contain many lecturers and a lower number of students (bottom right); these courses are to be found in the Faculty of Engineering & Built Environment (EBE) (“CHE4045Z,2019”, “EEE4022S,2019”) and in the Faculty of Health Sciences (FHS) (“IBS1007S, 2019”, “Molecular Medicine 2019”). That might be due to having outside industry-specific lecturers for the various courses.

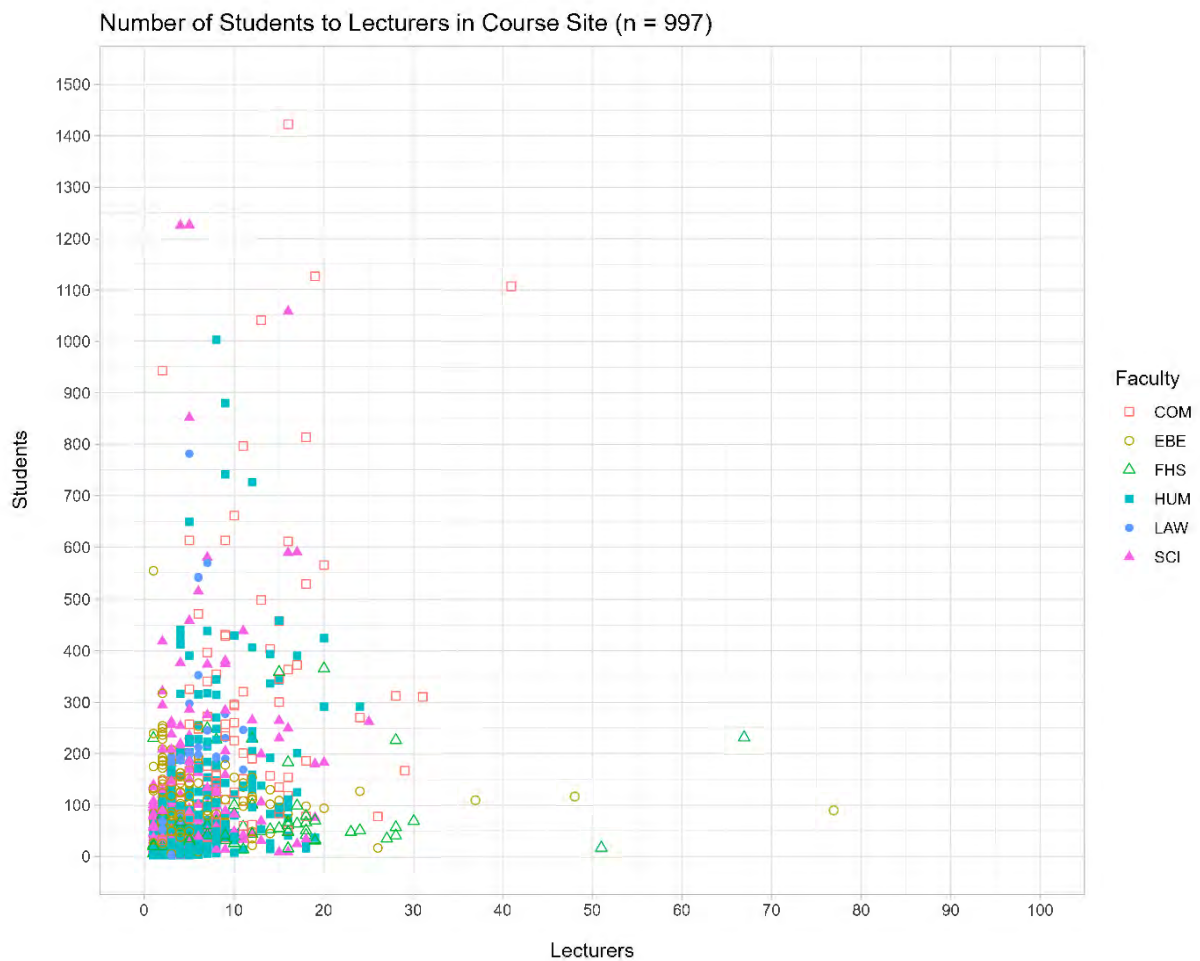


Figure 32: Number of Students to Lecturer in participants grouped by Faculty.

It maps the number of lecturers to the number of students, and Figure 33 shows the percentage of students and lecturers enrolled users on a site. It is evident that most users in the course sites are students, as the points are clustered around the top left side of the graph. There are several Humanities (HUM) courses around the (0.5, 0.4) mark as is a single course in the Faculty of Engineering & Built Environment (EBE) (“EEE4022F,2019”) and Faculty of Health Sciences (FHS)(“Molecular Medicine 2019”), and then three courses from the Faculty of Science (SCI) (“STA2007H,2019”, “CSC2005Z,2019”, “CSC3023F,2019”).

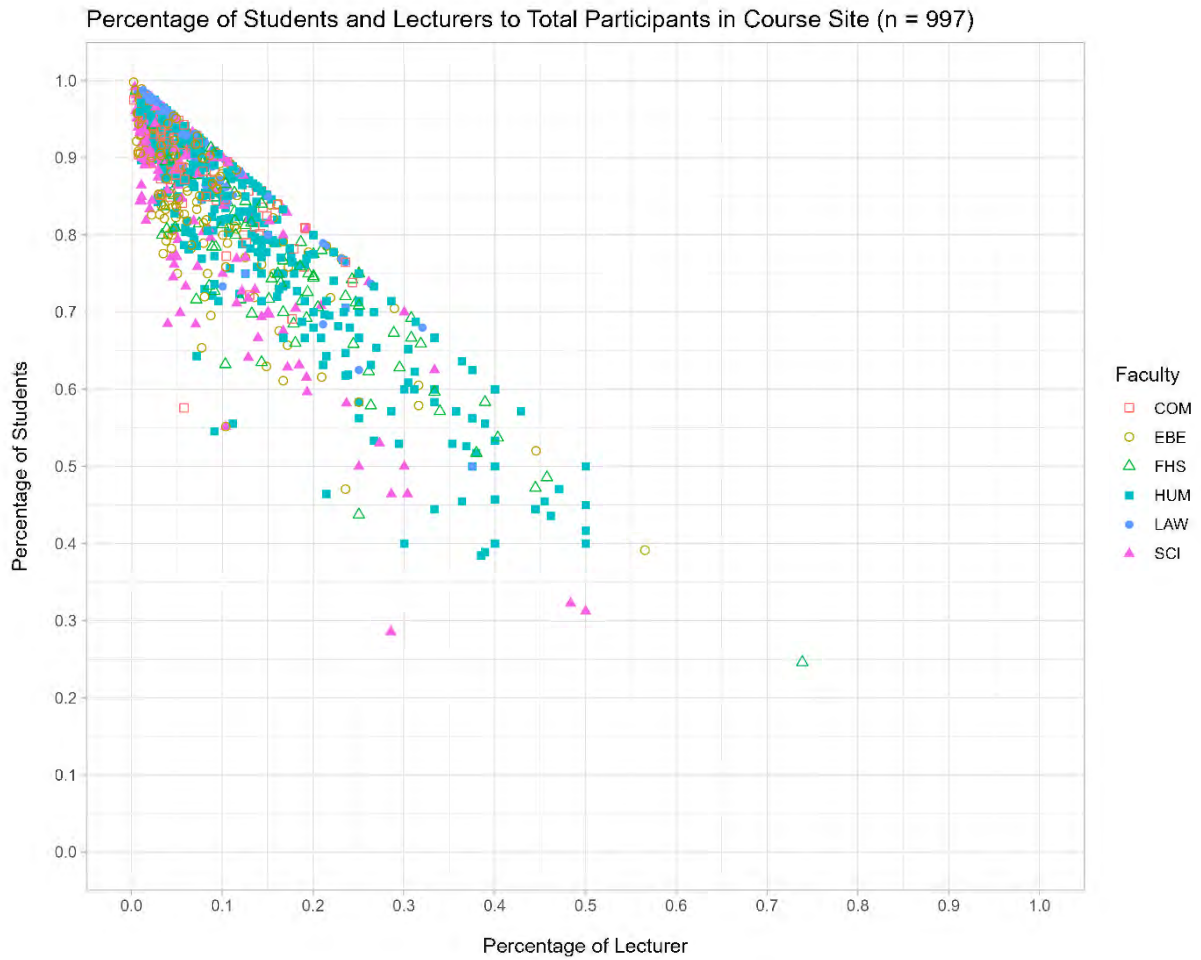


Figure 33: Percentage of Students and Lecturers in participants grouped by Faculty

Table 35 shows the enrolment details for those sites:

Faculty	Title	Lecturer	Lecturer %	Student	Student %	Support Staff	Tutor	Other	Total
EBE	EEE4022F,2019	26	56.52%	18	39.13%	2	0	0	46
FHS	Molecular Medicine 2019	51	73.91%	17	24.64%	1	0	0	69
HUM	German IIIB, 2019	5	38.46%	5	38.46%	0	0	3	13
HUM	Portuguese IIA 2019	5	50.00%	4	40.00%	0	0	1	10
HUM	REL2054F,2019	8	40.00%	8	40.00%	0	1	3	20
HUM	SLL3122S,2019	3	30.00%	4	40.00%	0	1	2	10
HUM	Spanish IIIA 2019	7	38.89%	7	38.89%	0	0	4	18
SCI	CSC2005Z,2019	16	50.00%	10	31.25%	6	0	0	32
SCI	CSC3023F,2019	15	48.39%	10	32.26%	6	0	0	31
SCI	STA2007H,2019	4	28.57%	4	28.57%	6	0	0	14

Table 35: The lowest percentage of students and lecturers per course site

The next section looks at the data extraction process to get the content for the course sites and the user interactions with the tools.

4.1.8 Sakai Database Structure – Site and Tool Tables

This diagram displays the table relationships between the tools available in Sakai and the course site table (middle top).

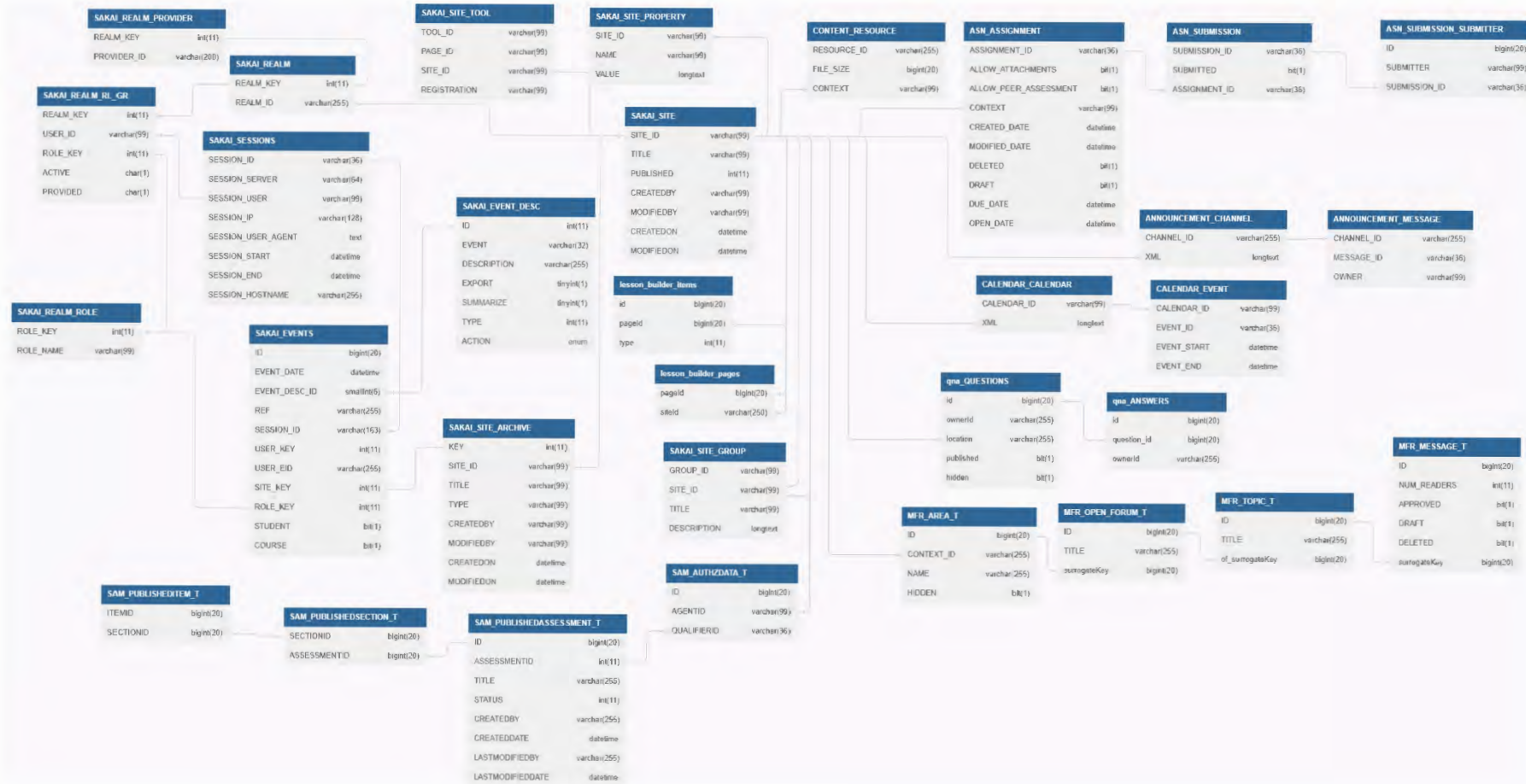


Figure 34: Relationship between the site table and all the tool tables.

Each tool in Sakai has a set of tables to describe the information used in the tool and a relational reference to the site (SITE_ID). The data was exported as CSV files with the selection of important fields that include creation information, ownership, and details of user activity (this varies by tool).

There are two main content type tools used in course sites in the LMS, namely “Resources” and “Lessons”.

The “Resources” tool allows instructors to share a wide variety of files with their students within a site (individual users may also have resources within their personal “My Workspace” area). The “Resource” query lists all the files in a course site and contains the SITE_ID, filename, size, and type (mime type). This information only pertains to files. It does not include permission and visibility information inherited by folders. This file visibility for students is remedied by using the advanced data export methods described later in this study.

Users with the correct permissions can upload almost any file (might be restricted by file size, depending on configuration) that they want to make available to other users on the site. These files can include documents, spreadsheets, slide presentations, audio, and videos, as well as web pages (HTML), simple text documents, library citations, and shared links to useful websites. The files and links can be organised into folders and subfolders, making it easier for users to locate and access. The file entry can include contextual remarks, which can be shown during specific dates and times. Lecturers can notify students by email when an item is added to the resources (*What Is the Resources Tool?*, 2023).

Whenever a file gets uploaded to resources, the metadata for the file gets updated in the table (database). The metadata contains, among other things, the name of the file, the size of the file, and, important for our comparison, the type of file.

The type, also known as the Multipurpose Internet Mail Extensions (MIME) type, indicates what kind of file this is. MIME types are defined and standardised in IETF's RFC 6838 (Freed et al., 2015).

The MIME type structure consists of 2 parts: a *type* and a *subtype*, separated by a slash (/), with no whitespace in between. Each kind of file will have a combination of *type* and *subtype*, always both, never just one.

type/subtype

The *type* indicates the general category of the data, such as audio, video, or text.

The *subtype* identifies the exact kind of data; for example, “video/mp4” denotes that this is a video file in the format of an mp4. Another example is “application/pdf”, which shows that the file is associated with a specific application (pdf - Portable Document Format). To make the comparison with “Lessons” easier, some “type” for files will be changed to a new factor called “document”, which describes reading or study material. These are text/html, text/plain, application/msword, application/epub+zip, application/pdf, text/x-r-markdown, text/x-r-markdown, etc.

Figure 35 indicates the number of discrete types of files, grouped by main “type”, followed by a graph which shows the total number of that “type” and, finally, the total size of that “type” in the data set.

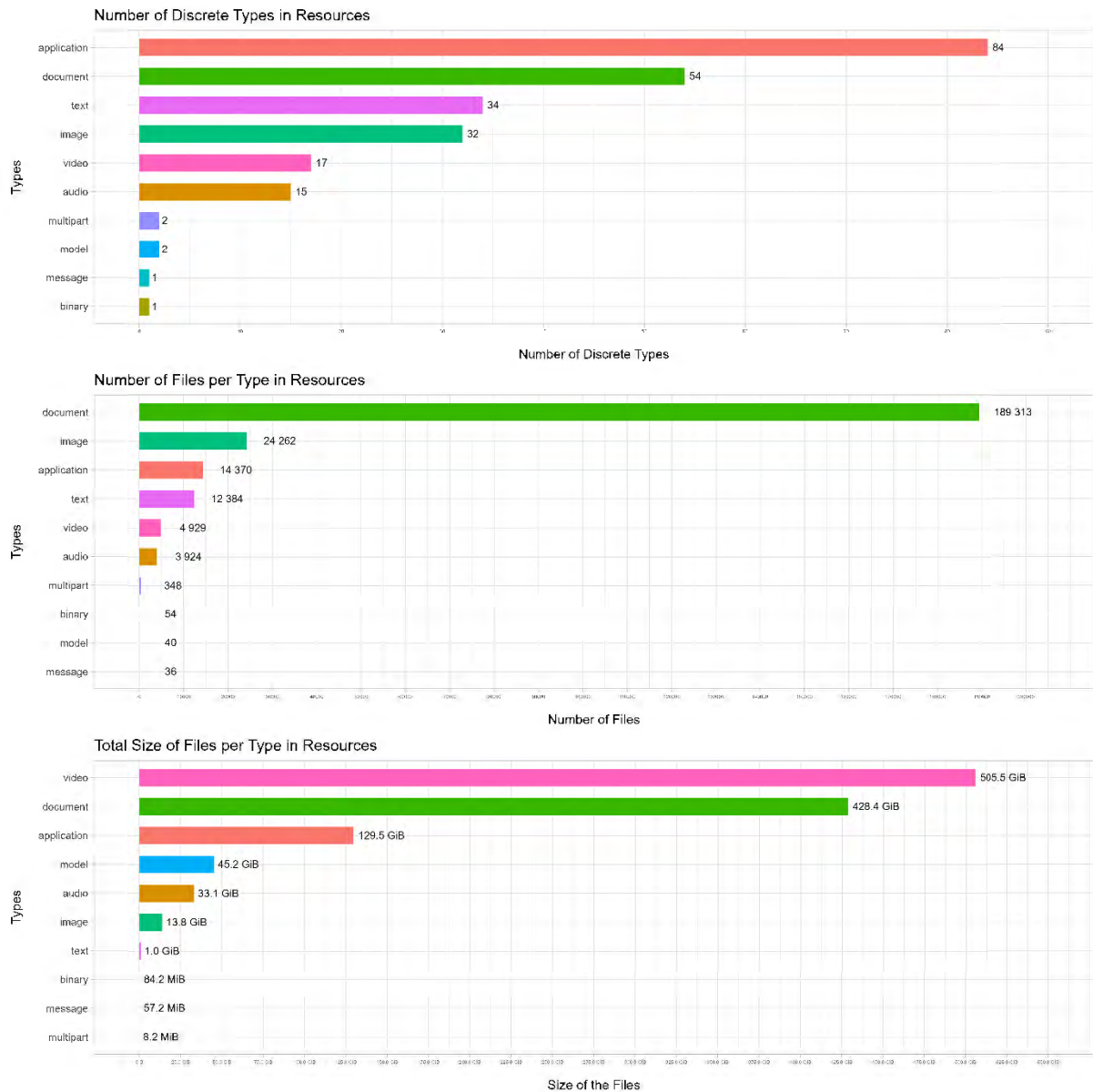


Figure 35: Number of discrete mime types, number of files in resources, and size of the files in resources.

The “Lessons” tool allows lecturers to structure the learning pathway for students by constructing pages with various items (item types are shown below) and linked to the navigation bar (*What Is the Lessons Tool?*, 2023). The items can be arranged in any order and give a high level of flexibility to the course designer (lecturer) in terms of scaffolding the learning experience from a content delivery perspective.

- a) Text item (i.e., content on the page – Rich Text / HTML) (type:5)
- b) Content links to items in Resources such as files (type:1) or URLs (type:6)
- c) Link to a Resource folder (type:20)
- d) Multimedia Link (like Video) (type:7)
- e) Links to published assignments (type:3) or assessments (type:4)
- f) In-line question items on the page (type:11)
- g) Links to forum topics (type:8)
- h) Links to subpages (type:2)
- i) Checklist (type:15)

- j) Published Announcement (type:17)
- k) Learning Tool (LTI) (type:12)
- l) Student comments on the page (type:9)
- m) Calendar Entry (type:19)
- n) Links to student pages where students may create content (type:10)

As with “Resources”, the “Lessons” tool pages can be analysed, and the items in the pages can be grouped by their type. Some of the types are structural and not relative to content comparison, e.g., Break or Section (type:14). There are some, like the Resource Folder (type:20), which will be excluded as it already includes the resource file information in another part of the data set and it unnecessary to re-introduce the files here.

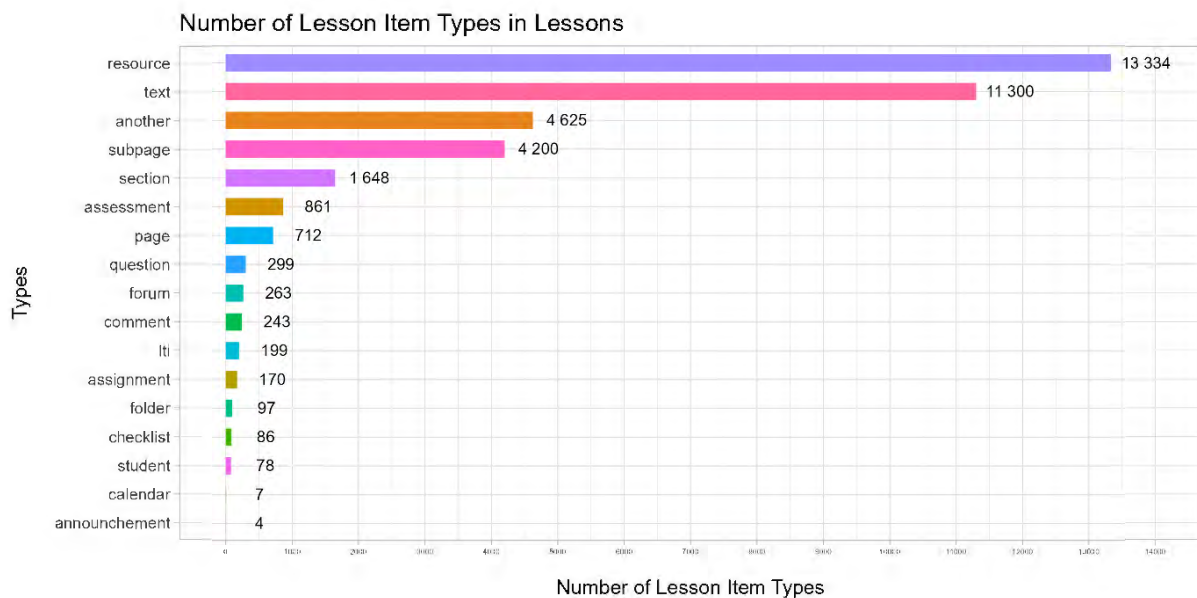


Figure 36: Total Number of Lesson Types

4.1.9 Advanced Data Export

At this stage, the data set contains the site information, providers, participants, and content provided by the tools. The tools in the course data set, because it was extracted directly from the database, give an administrator view of the site structure (which is similar to “Lecturer” / “Site Owner”) to be able to evaluate student interactions with the site, the student view of the site is also required to evaluate the impact of the tools on their learning experience. The visibility of the tool in a course site is defined by the role associated with the logged-in user and the permission structure of the realm associated with the site. Constructing a database query to extract this information would be too complex to evaluate. The logic of which tools are visible in the navigation bar is part of the Sakai codebase.

Another issue that was found is that the information about linked tools, which use Learning Tools Interoperability (LTI), and outside web pages were incomplete, as it does not show the URL or properties for the link, which means that they could not be distinguished from each other (except through title, but that is not a dependable identification method).

The access that a user has to interact with a tool in the course site changes depending on their assigned role. As an example, the “Lecturer” role can create pages, add content links to resources and define the flow of lesson content, whereas the “Student” role can access the page and read the

pages but will not be able to change the content. Similarly, In the Assignment tool, the “Lecturer” can create and change assignments, and the “Student” can submit assignments, which might include uploading files. The “Lecturer” role can view all the submitted assignments and files, whereas a “Student” role will only have access to their content, and a “Tutor” might be able to view all the submissions for a group of students.

The tool visibility information (permissions for the tool) and the configuration for the linked tools are obtainable when using the Sakai API. An Application Programming Interface (API) allows the access to internal logic to access, update, or delete data in a system. In this case, Sakai has a Simple Object Access Protocol (SOAP) and a REpresentational State Transfer (REST) interface ([server address]/sakai-ws/). To retrieve the site information (tools, structure, permissions, and configurations), the following steps are required: authorising the call with Sakai, requesting the data from a specific address (also called an endpoint), and processing the data received. This is similar to the process diagrammed in “Students' behaviour analysis under the Sakai LMS”, Fig.3 Extract roster and the data of assignments and tests (page 252, Wan et al., 2018). Using the SOAP API, the request response is returned as a well-formed XML document. In this study, the REST API is used to obtain the data, and the response object is formatted as JSON (JavaScript Object Notation). The site description for each site is obtained, then the tools (called pages in navigation context) and the permissions for the site. This resulted in 7,527 files (3 files per site – site details, pages, permissions; for 2,509 sites). The details, pages, and permissions were combined into a comprehensive data set that defines the visibility of each tool per role, then processed to create a data set for all the tools on the site and one for tools visible to students.

4.1.10 Accessing Outside Data Sources

IMS Learning Tools Interoperability (LTI) is a standard describing the details of the connection between third-party learning applications (a learning tool) and the learning management system (LMS). The learning tools generally add additional functionality that is likely not part of the standard LMS experience. There are a variety of tools available, including eBook readers, attendance, geographic data representation, Turnitin, and lecture recording.

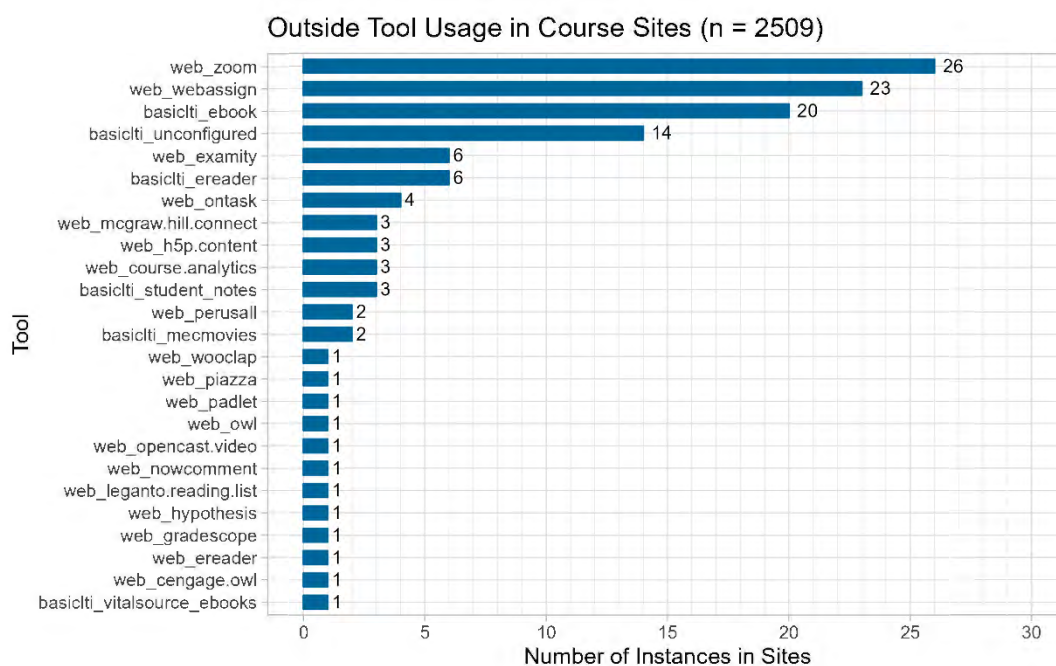


Figure 37: Number of tools in All the course sites

Figure 37 sets out the total number of outside linked tools in all the 2019 course sites; from this representation, the top 5 tools are a web link to Zoom (an online video collaborative meeting service), WebAssign (a Learning platform providing assignments and other instructional tools), an LTI link to a generic e-book reader, and an LTI link that has not been configured at all. Figure 38 uses the subset of course sites and shows the number of tools in each site, visible to lecturers (all) and the tools visible to students.

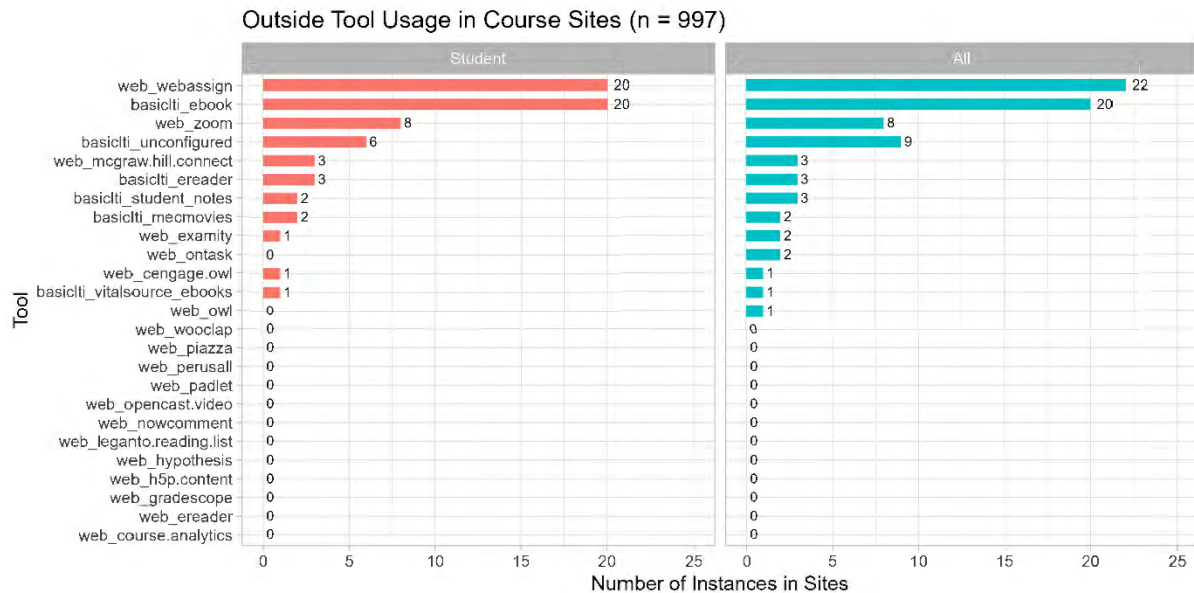


Figure 38: Number of tool instances in course sites grouped by visibility to students.

The same list of available tools are shown, but it is obvious that some tools are not used in undergraduate sites (count is 0). Also visible are two local Tsugi tools (LTI Tool Framework - Severance, 2023), namely “Mec Movies” (Parker & Philpot, 2023) and “Student Notes” (Tsugi - Student Notes, 2023).

The most used LTI connection is the lecture recording system used by the University, which is called Opencast. Opencast is a scalable open-source video capture, management, and distribution system for academic institutions (Opencast, 2023).

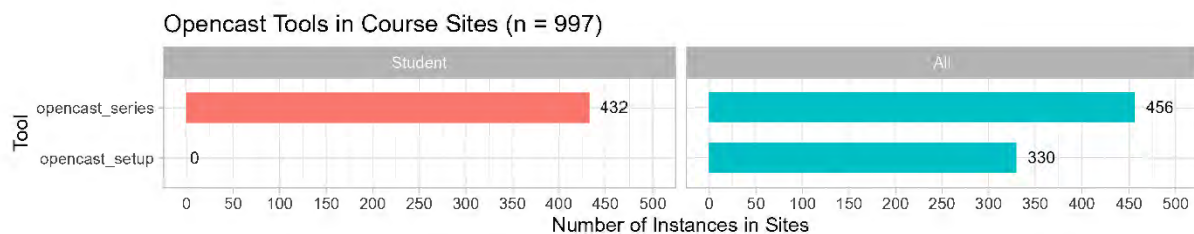


Figure 39: Number of Opencast LTI instances in course sites

Figure 39 indicates that close to half of all courses use Lecture recording, and almost a third (330) of course sites did not decide to use this tool and left it in the setup phase. There are some course sites (24) that have decided to use lecture recordings but have not made those recordings available to students. When completing the setup phase, the site is linked to a collection of videos (named a “series” in Opencast), and the unique identifier (series_id) to the collection is stored in the configuration of the site tool. The collection of videos (named “events” in Opencast) is associated

with the series, and the interaction information for users is stored in the database (oc_user_action, oc_user_session). The structure of which is shown below.

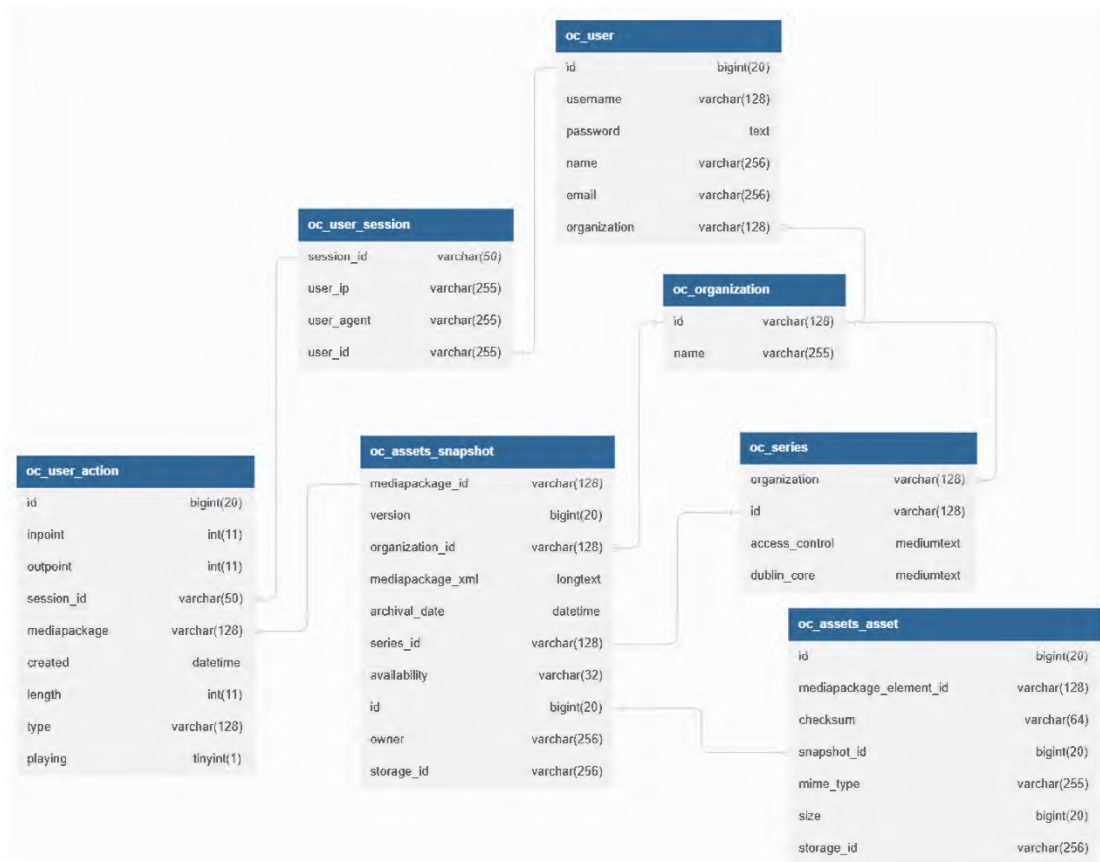


Figure 40: Opencast database structure.

Also included in this list of tools are tools that use the Web Content tool to add links to URLs (websites). These outside tools store student and lecturer interactions, which are not accessible without a specific request to each company for each linked tool. As such, it is beyond the scope of this analysis to obtain.

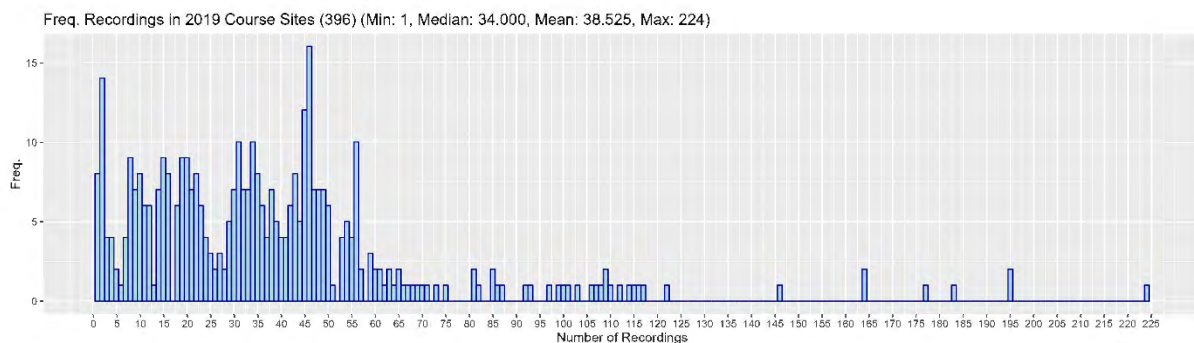


Figure 41: Number of recordings for Opencast Series

There are 558 series linked to all the 2019 course sites, of which 424 fit into the filtered data set. There are 28 collections (series) which do not contain any videos (events), leaving 396 series for which the number of videos (events) was plotted on a histogram above (Figure 41). This reveals that most sites contain around 34 recordings per series or roughly one recording for each teaching week of 2019.

4.1.11 Exporting Event Data

Opencast also tracks the number of recordings a user might view or download. The number of events is shown per week and grouped per year.

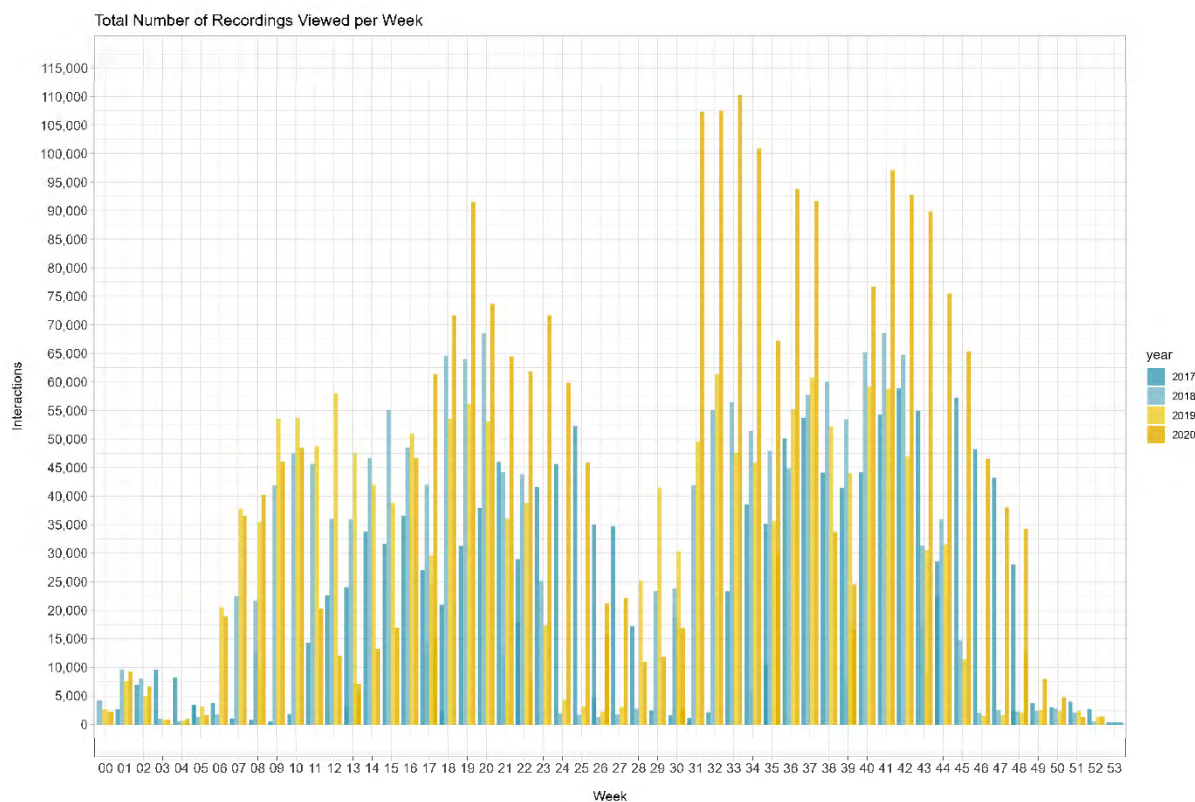


Figure 42: Opencast recording view events by week from 2017-2020

At the end of the analysis phase, after the clustering of the course sites by tools, the student interaction with the tools in the site will be used for evaluation. Sakai stores user interactions in two tables, namely SAKAI_SESSION and SAKAI_EVENT. As the names imply, the session table stores the overall information of a user’s access to Sakai. The user interaction with the course sites and tools is stored in the events table. On average, the interaction in the LMS system would equate to about ~23.7 million events per year (650K events per day).

The query exported all the events and sessions for the year (2019) grouped by day. This created 366 files (overlapping with 2018-12-31 to contain the session started before the year-end) as 9.83 GB of compressed CSV files.

Year	Events	Sessions
2017	236,925,475	19,965,054
2018	245,629,624	21,012,448
2019	236,951,152	23,444,574
2020	327,159,017	25,057,393

Table 36: Sakai total events and sessions per year (2017, 2018, 2019,2020)

Figure 43 takes the event data (summarised per year in the table above) for the years displayed (2017-2020) and plots it per week. The increases and decreases are indicative of the start and end of teaching periods, the quarterly structure, and vacation weeks (for more information, see Appendix – 2019 Academic Year).

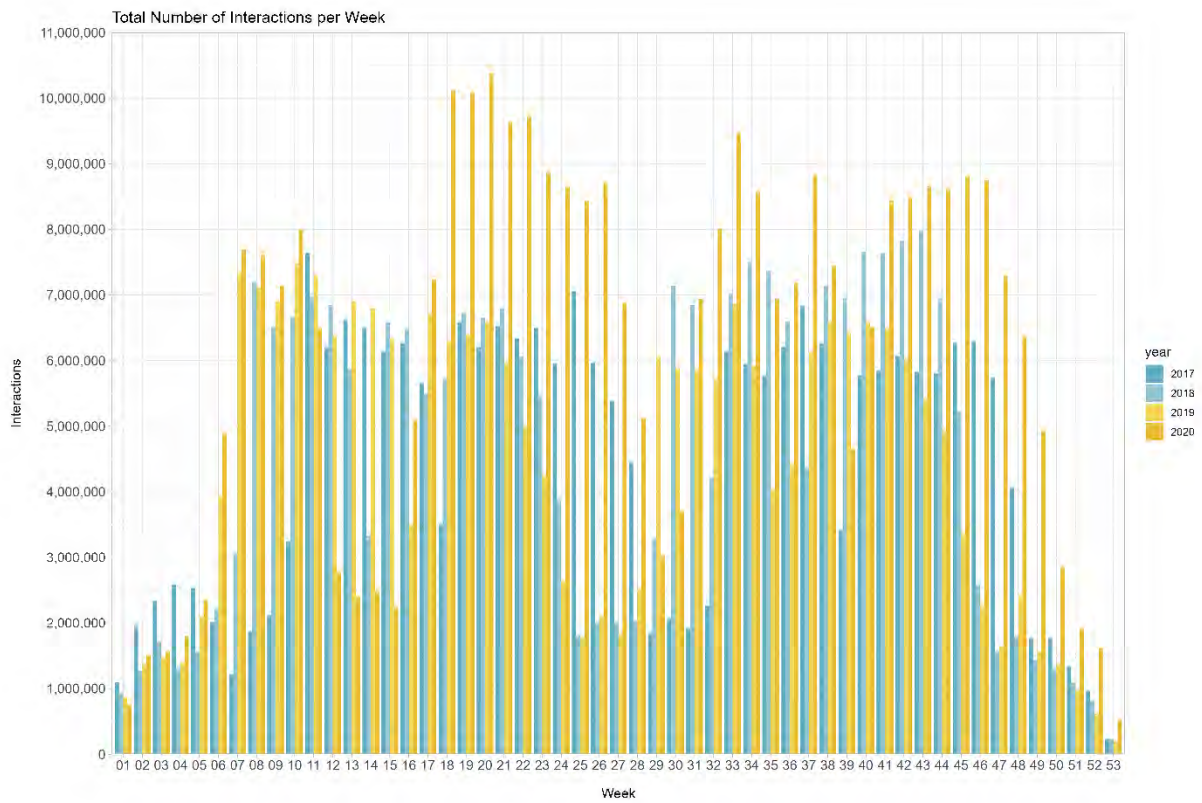


Figure 43: Number of user interactions per week for 2017-2020

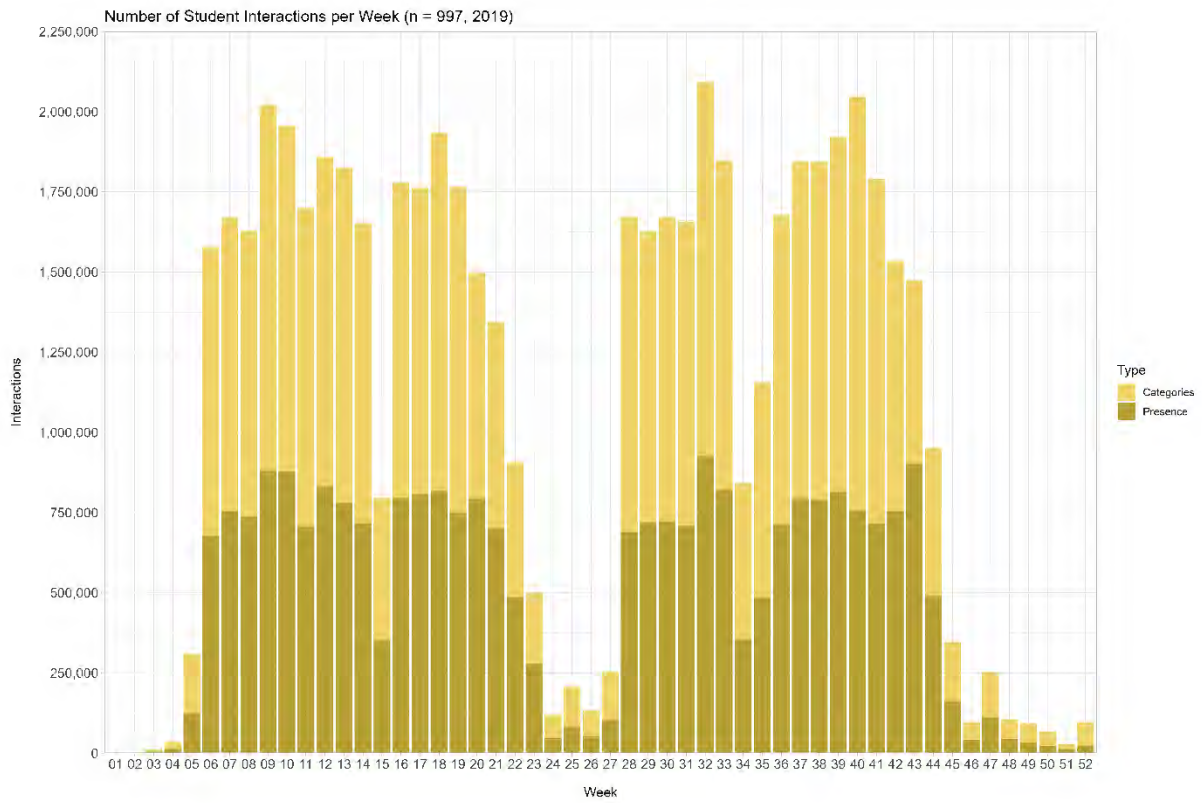


Figure 44: Number of user interactions per week on the filtered course sites for 2019.

Figure 44 plots the student interaction per week for the filtered course sites. The chart is split into two types: first, the events for user presence (accessing the site) and then the other part for all the interactions with the tools. The events graph gives some scope as to the number of events which will be later used in the site classification to map the accessed tools and content.

4.1.12 Summary

This section started with a list of all possible courses (5,209) from the Student Information System (SIS). This does not mean that all the courses have corresponding sites in the LMS, as sites might have multiple courses linked to them or because that course type might not be offered in that year.

The initial data set, obtained from Sakai - the LMS of UCT, contained details for 2,509 course sites. During the extraction and validation process described in this section, the course site data set was filtered according to the following parameters:

- I. Course site has a valid Faculty, Department, and Academic Career Code.
- II. Academic Career Code is "UGRD", "HONS+UGRD", "MEDS" (Undergraduate or Medical Degree).
- III. It is an undergraduate course (filtered by the year of course code).
- IV. The course site is not for "MBChB" (which is a content site for a collection of different course codes).
- V. Term is 2019.
- VI. Contains more than three students.

Applying the constraints returns 997 course sites that will be used for the analysis phase. Additional permission data was obtained using a more advanced extraction process (REST API) so that the tool visibility and external tool linkage information could be added to the data set. The content of the sites was extracted and investigated to make it possible to cluster the sites based on content available to students, which included the lessons, resources, and lecture recordings. Lastly, the user interactions with the LMS were explored to indicate where and how the interactions took place and to inform on the scope of the data involved.

4.2 Processing Data

4.2.1 Data Set and Processing Steps

The course site data set (997 observations) contains variables to identify the site (title, unique identifier), the site details (creation date, modified date, published status), provider details (in section 4.1.5), tool details (all the tools and tools visible to students), content for each tool, lecture recording information (series and events), and finally user enrolment information (site participants).

The data set is split into participants (I), tools visible to students (II), and categories of tools visible to students (III). This will make it easier to distinguish the impact of each part and might give some insight into the tool choices for each course.

The first part (I) contains the participant information (as discussed previously), which describes the "Lecturer", "Student", "Tutor", "Support Staff", "Librarian", "Observer", and "Reviewer" roles. The permissions and focus of each can be divided into "Teaching" for "Lecturer", "Learning" for "Student", "Teaching Support" for "Tutor" and "Support Staff", "Other" for "Librarian", "Observer" and "Reviewer". The permissions for the roles reflect this focus. The "Lecturer" role has full access to tools and content. The "Student" role is focused on interactions with the content. "Teaching Support" has limited access to add content but has a larger view of student activity as the role is designed to support the lecturer in teaching and learning. Lastly, the "Other" roles are restricted to primarily read-only access.

The next part (II) is the analysis of all the tools in conjunction with the tools visible to students. The tools selected for the course sites can be categorised into the four main categories identified in the literature review (part III). This is to reduce the number of variables and possibly improve the clustering results but also to make the results easier to compare to other LMSs.

Each part will be analysed to determine if it does have clustering tendencies, using a statistical method (Hopkins statistic) and a visual assessment of tendency (VAT). The Hopkins statistic (Lawson & Jurs, 1990) is used to assess the clustering tendency of a data set by measuring the probability that a given data set is generated by a uniform data distribution (Kassambara, 2023b). The Hopkins statistic value ranges between 0 and 1; if the value is close to 0 - that indicates the data set is uniformly distributed and lacks cluster structure; else, if the value is close to 1, then that suggests the data set has a significant cluster structure, and it's suitable for clustering analysis.

The visual assessment of cluster tendency (VAT) (Bezdek & Hathaway, 2002) is used to visually examine a data set's clustering tendency or inherent structure. It helps identify whether the data exhibits a clear pattern of clusters or if it appears more random or uniformly distributed. The image it produces can be inspected for blocks or patterns indicating clusters within the data set.

4.2.2 Course Participants

The participant data set contains 997 observations and 7 general roles in 4 groups.

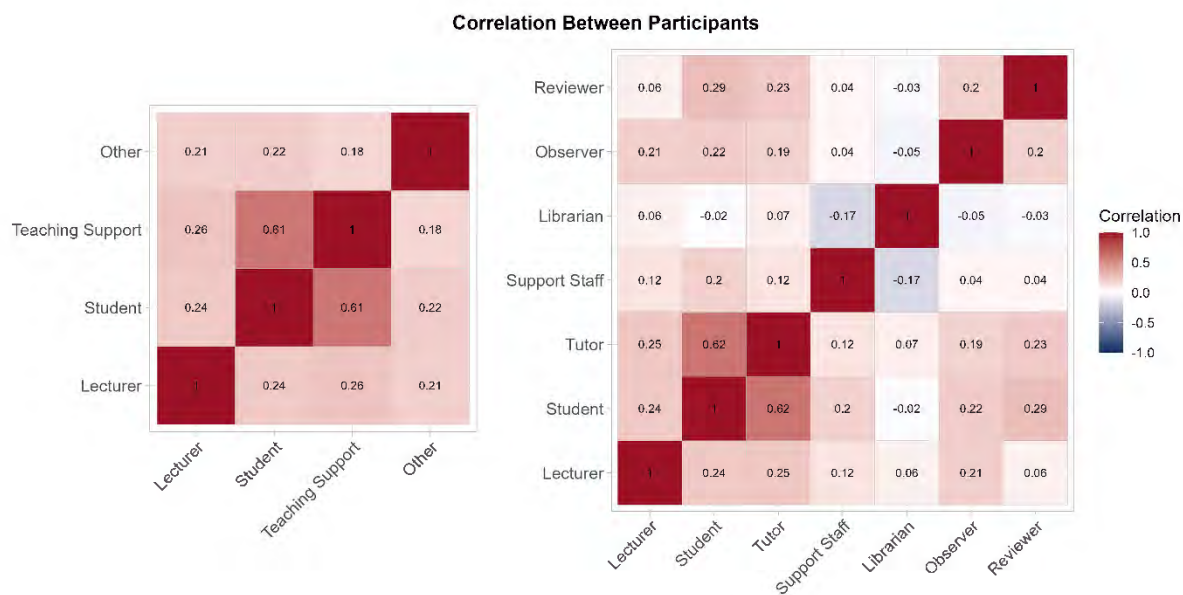


Figure 45: Correlation matrix of course participants

Figure 45 shows two correlation matrix graphs. The first (left) is for the participants combined into the four categories, which is followed by the correlation matrix of all the roles on the right. There is generally a positive correlation between all the roles except for the relationship between “Librarian” and “Support Staff”, which has a weak negative correlation (-0.17). “Student” and “Tutor” have the largest correlation of 0.62, and this significance is translated into the four categories graph where “Teaching Support” (a combination of “Tutors” and “Support Staff”) has a moderate positive correlation of 0.61.

Lecturer	Student	Tutor	Support Staff	Librarian	Observer	Reviewer	Total
6,545 (5.16%)	114,009 (89.85%)	2,898 (2.28%)	1,952 (1.54%)	312 (0.25%)	1,160 (0.91%)	10 (0.01%)	126,886 (100.00%)

Table 37: Total participants for 7 roles in the course data set (997)

As Table 37 shows, in total, the “Librarian” users are 0.25% of the total participants, compared to the 1.54% of “Support Staff”; this weak negative correlation can be safely ignored. This also shows that the majority (89.85%) of enrolled users are students (confirmed in Table 38); it is therefore likely that the clustering on participants will centre around the number of students per course site.

Lecturer	Student	Teaching Support	Other	Total
6,545 (5.16%)	114,009 (89.85%)	4,850 (3.82%)	1,482 (1.17%)	126,886 (100.00%)

Table 38: Total participants for 4 combined roles in course data set (997)

According to the Hopkins statistic (shown in Table 39), both data sets have a high degree of clusterability as both values are approaching 1. While there is not much difference in the Hopkins statistic between the two data sets, the VAT image for the summarised role data set (shown in Figure 46) does reveal a more distinct cluster structure. Using the summarised role data set reduces the number of variables and improves comparability with other LMSs with a similar role structure.

Summarised Roles	All Roles
\$hopkins_stat [1] 0.96889	\$hopkins_stat [1] 0.97882

Table 39: Hopkins statistics for the participant's data set.

The graph in Figure 46 displays the visualisation of the distances between the site participants. The colour value indicates the proportion dissimilarity between observations, meaning darker colours if $\text{dist}(x_i, x_j) = 0$ and white if $\text{dist}(x_i, x_j) = 1$. Observations belonging to the same cluster are in consecutive order. Using this graphing method, it's possible to distinguish a number of large blocks on the diagonal, which indicates that the data set is highly clusterable.

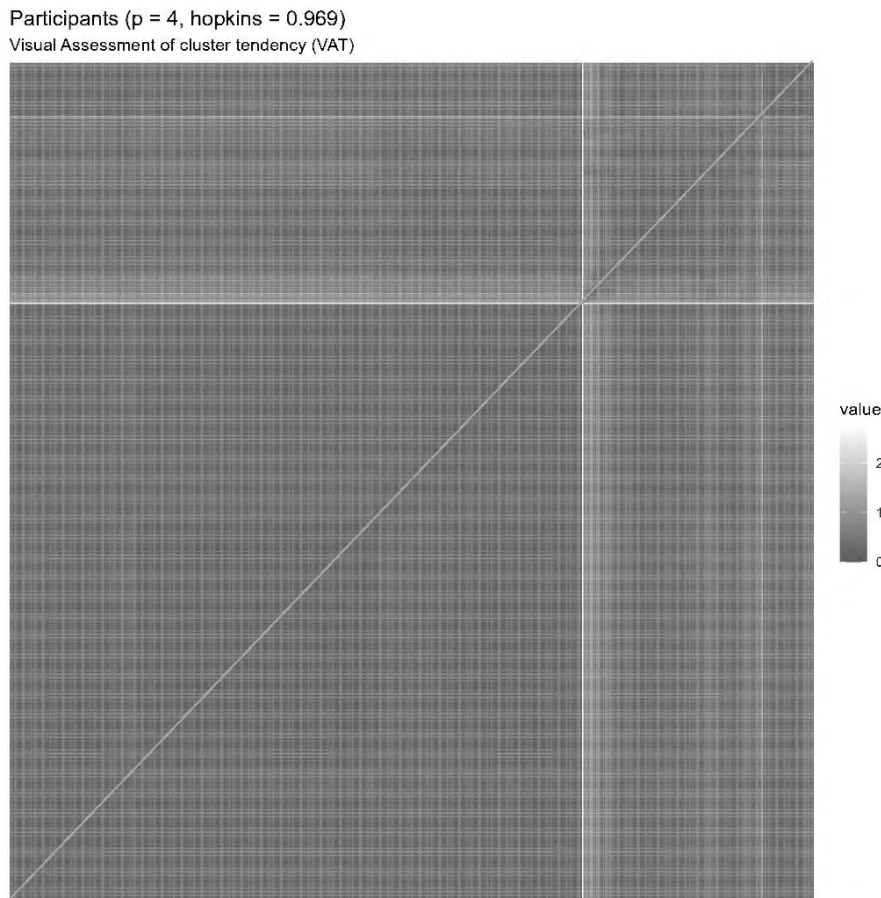


Figure 46: Visual assessment of cluster tendency (VAT) of summarised roles

4.2.2.1 K-Means

Firstly, deciding the optimal number of clusters to use is vital. This can be calculated by comparing the different number of clusters (k) and comparing the gap statistics, the total within sum of squares, and the silhouette method.

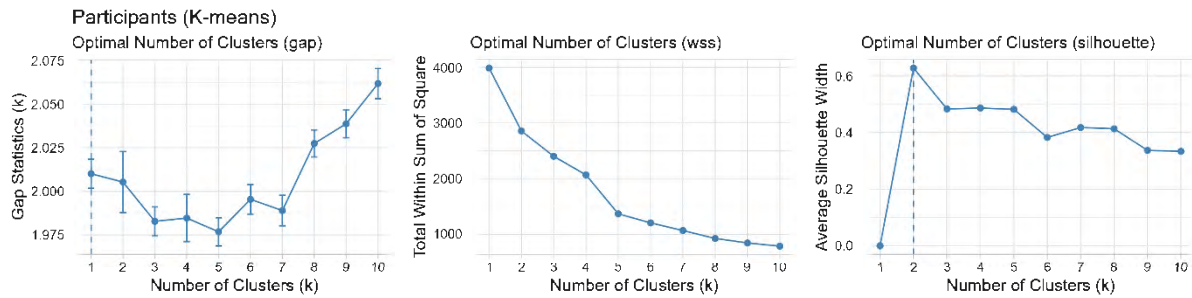


Figure 47: K-means – Determine the optimal number of clusters using (gap, wss, silhouette) for participants

Figure 47 shows the graphs for “Optimal Number of Clusters” using three methods – Gap analysis, total within sum of squares (WSS), and Silhouette analysis.

The first (leftmost [gap]) is the Gap statistics; this compares the within-cluster dispersion of the observations to that of a reference random distribution. The gap statistic is plotted against the number of clusters (k); smaller values (y-axis) suggest that the current number of clusters is not significantly better than the random distribution, and larger values indicate that the current number of clusters is substantially better than the random distribution. The ideal graph would go from a low gap value and a low number of clusters, increase to higher gap values, and decrease as the number of clusters increases. This graph is almost the opposite of that description as it starts high, drops to a low gap value with a low number of clusters, drops lower as clusters increase and then increases in the latter part of the plot. The optimal number of clusters indicated is 1, which means that either K-means is not the best clustering method, or this measure is non-determinative.

For the total within sum of squares (WSS) plot (in the middle [wss]), the number of clusters is indicated by the bend or elbow in the line. The WSS value decreases (y-axis), and the number of clusters increases (x-axis). The WSS decreases to zero as the number of clusters increases. The elbow (bend) is where additional clusters give diminishing returns and indicate the lowest number of clusters with a low WSS value. The line indicates that either two or five would be a good choice for the number of clusters, after which it does level out.

On the right (silhouette) is the silhouette method; this plots the quality and consistency of the clusters by quantifying how well an observation (o) fits within its assigned cluster comparatively to another cluster. To calculate this, take (o) and calculate the average dissimilarity to observations within its own cluster - a measure of cohesion in the cluster, call this (a). Calculate the average dissimilarity between (o) and all observations in its nearest neighbouring cluster – a measure of separation to the closest cluster, call this (b). To calculate the silhouette coefficient (s) for the observation, the difference between (a) and (b) is divided by the greater of the two.

$$s = (b - a) / \max(a, b)$$

The average silhouette coefficient across all observations measures the cluster quality. If (a) is small, then cohesion is good (small distance between elements in their own cluster), and if (b) is large, then there is good separation between the nearby clusters. In the graph, this shows up as a large value, which indicates that two is the optimal number of clusters; it drops slightly for three to five clusters and then lowers as the number of clusters increases.

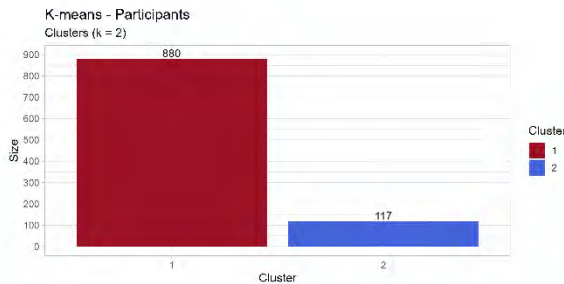


Figure 48: K-means (k=2) size of participant clusters

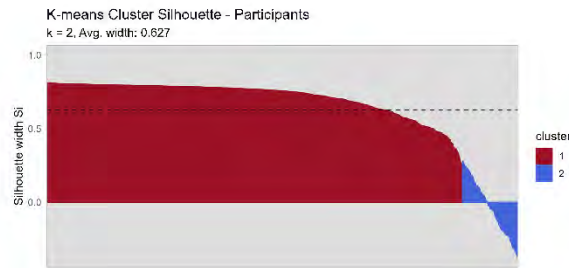


Figure 49: K-means (k=2) cluster silhouette of participants

Figure 48 presents the number of observations in each cluster when (k=2). The silhouette graph (Figure 49) suggests that two clusters might not be the right choice. Ideally, the average silhouette height (indicated by the dashed line) should intersect with all the clusters. The silhouette score measures how well the data is clustered. A score closer to 1 means the data is perfectly clustered, while a score of -1 indicates that it doesn't belong to any cluster. A score of 0.0 means the point is on the edge of two clusters. If the average silhouette score is close to 0, it indicates poor clustering. This could be due to the data not being clusterable or an incorrect number of clusters (k) was chosen. In the silhouette graph (Figure 49), the average silhouette score is above 0.5, indicating a reasonably good clusterability. However, more than half of the second cluster is below 0.0, which most likely suggests that those observations overlap with the other cluster.

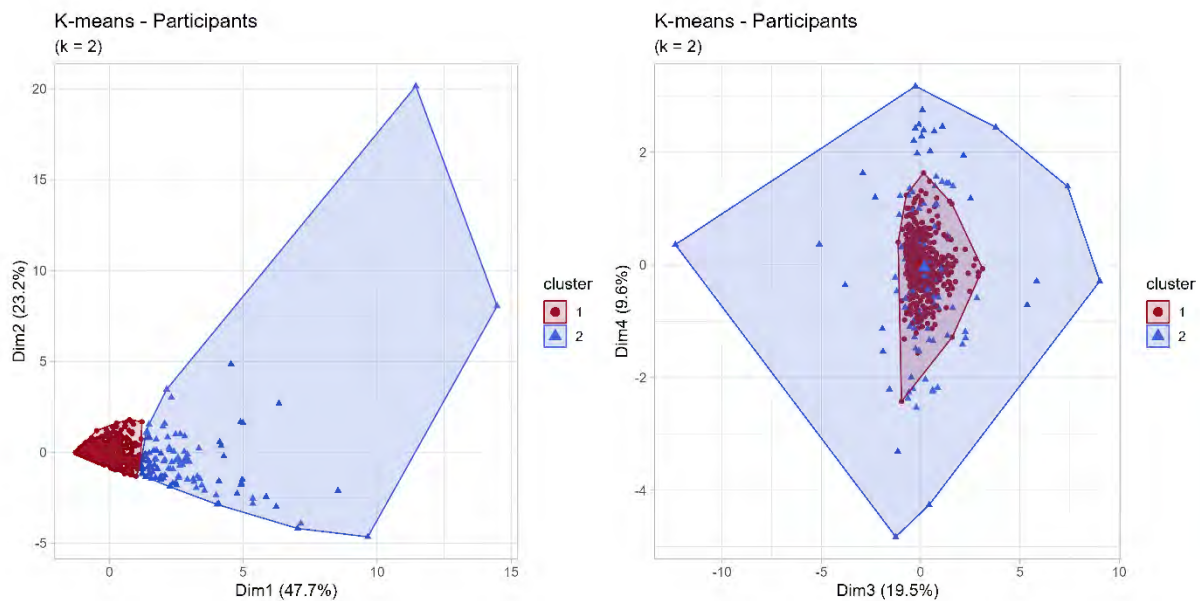


Figure 50: K-means clustering of participants (k=2)

Taking the mean values of the two clusters based on the summarised roles, it's evident that the first cluster contains sites with few participants, and the second cluster contains more participants, mostly comprised of "Students".

Cluster	Lecturer	Student	Teaching Support	Other
1	1 : 5.616 : 30 (4.367)	4 : 74.945 : 554 (70.581)	0 : 3.173 : 18 (3.596)	0 : 0.882 : 16 (1.430)
2	2 : 13.701 : 77 (11.672)	17 : 410.744 : 1423 (279.106)	0 : 17.590 : 52 (11.485)	0 : 6.034 : 195 (20.949)

Table 40: K-means (k=2) min:mean:max (standard deviation) of participants in summarised roles

Table 40 details the two clusters, broken down into the minimum, mean, maximum and standard deviation for each summarised role in the participant data set. It indicates that the mean for each

role is higher in the second cluster, as is the minimum and maximum. This reinforces the findings in the histogram of students in course sites (Figure 23), which indicated that the majority of sites have a low number of students (less than 100), which ended up in cluster “1”, and a few sites with a higher student count, which are represented in cluster “2”.

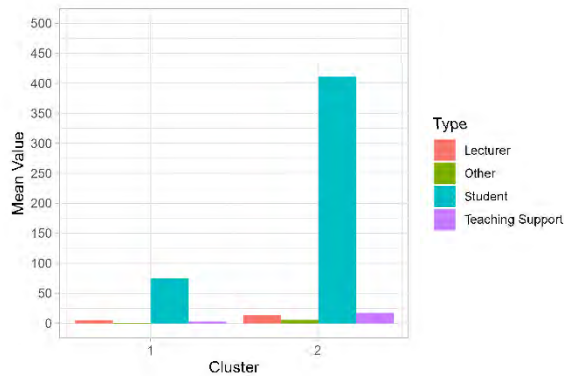


Figure 51: K-means (k=2) mean of roles

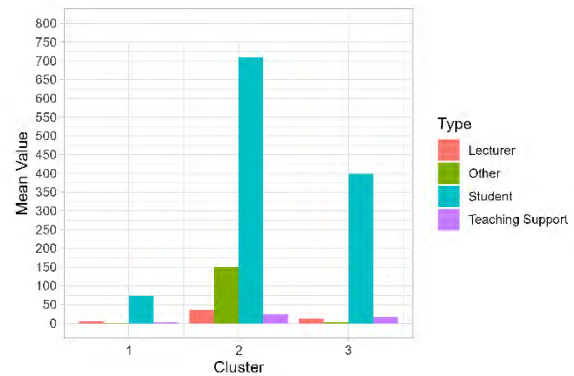


Figure 52: K-means (k=3) mean of roles

Comparing the mean value of each role between k=2 (In Figure 51) and k=3 (In Figure 52), it’s possible to differentiate the clusters into large, medium and small numbers of participants. However, this conclusion is not supported by the size of the clusters (Figure 53), which should ideally be relatively equal, or the silhouette graph (Figure 54).

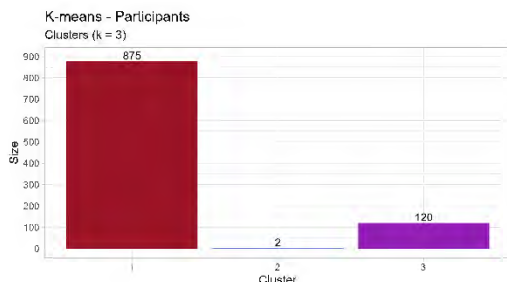


Figure 53: K-means (k=3) size of participant clusters

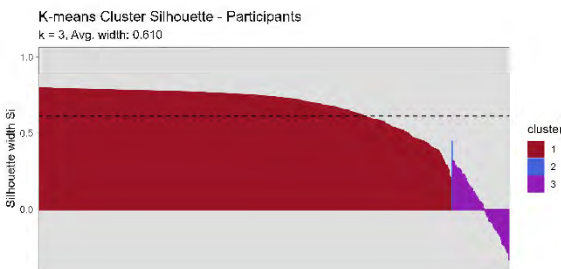


Figure 54: K-means (k=3) cluster silhouette of participants

Using 2 clusters (i.e., k=2) produce the best result for the participant summarised data set and that the division is largely based on the number of students in each course site.

4.2.2.2 K-Medoids Clustering – Partitioning Around Medoids (PAM)

In a similar approach as with K-means, the optimal number of clusters can be determined by looking at the gap statistics, the total within sum of squares, and the silhouette method (Figure 55).

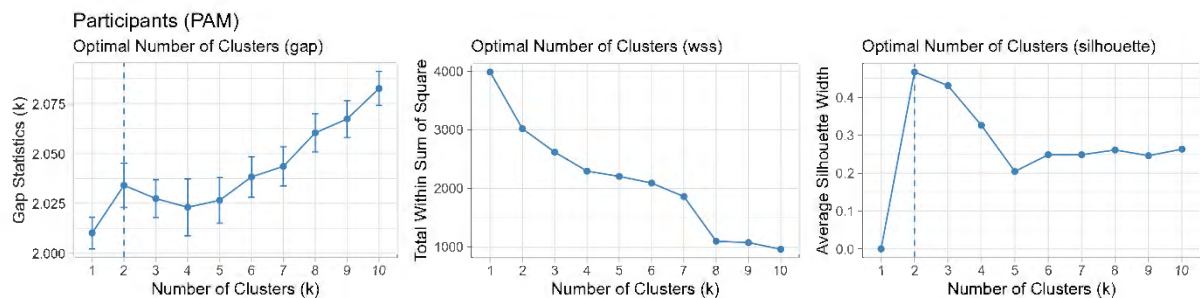


Figure 55: PAM – Determine the optimal number of clusters using (gap, wss, silhouette) for participants

The elbow method, the middle graph in Figure 55, suggests that either four or eight clusters might be a good choice, while the gap statistic (left) and silhouette method (right) both suggest two clusters as optimal.

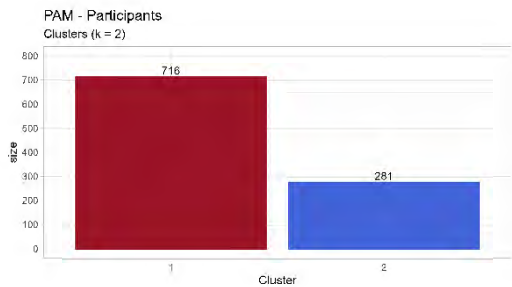


Figure 56: PAM (k=2) number of participants in clusters

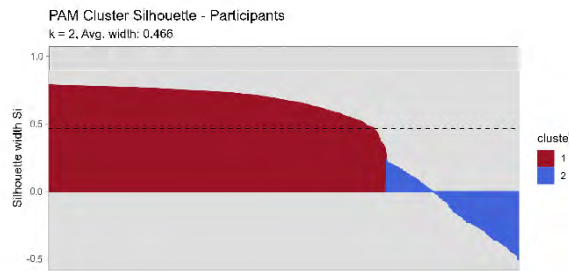


Figure 57: PAM (k=2) cluster silhouette of participants

The size of each cluster in Figure 56 reveals the observations more equally distributed than the K-Means clustering method for the same number of clusters (re-enforces that this algorithm is less sensitive to noise and outliers). The silhouette plot (Figure 57) still indicates that there might be many samples wrongly assigned. There are some observations below the 0.0 line. Like the clusters of K-Means, the PAM clusters also seem to be divided into sites with a low number of participants in cluster 1 and a high number of participants in the second cluster.

Cluster	Lecturer	Student	Teaching Support	Other
1	1 : 4.690 : 18 (3.061)	4 : 58.205 : 258 (53.686)	0 : 2.108 : 13 (2.295)	0 : 0.799 : 16 (1.342)
2	1 : 11.342 : 77 (9.198)	10 : 257.416 : 1423 (231.383)	0 : 11.890 : 52 (9.459)	0 : 3.238 : 195 (13.753)

Table 41: PAM (k=2) min:mean:max (standard deviation) of participants in summarised roles

There is a clustering algorithm, Clustering Large Applications (CLARA) (Kaufman & Rousseeuw, 1990), that is an extension of PAM; it can better handle large data sets and reduces the computing time needed (Davidson, 2022). The algorithm starts with randomly sampling a subset of the original data set. This subset is then used to generate the medoids of the clusters, after which the medoids are used for the entire data set - similar to PAM. Although it has the benefit of running faster than PAM, it is not suitable for a data set of about a thousand observations. The results were quite close to the existing PAM results without a significant decrease in computation time; as such, it was excluded from this study.

4.2.2.3 Hierarchical Clustering (AGNES & DIANA)

4.2.2.3.1 Agglomerative Nesting Hierarchical Clustering (AGNES)

As with other clustering methods, the gap statistic, total within sum of squares (WSS), and silhouette calculations can help determine the optimal number of clusters.

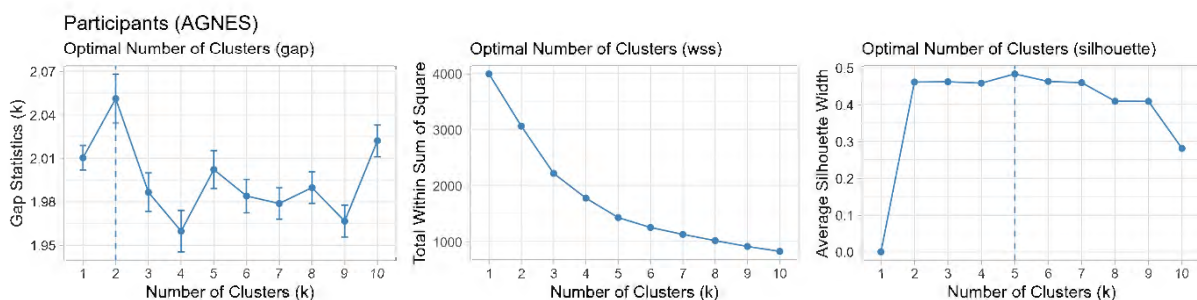


Figure 58: AGNES – Determine the optimal number of clusters using (gap, wss, silhouette) for participants

According to the gap statistical analysis in Figure 58 (left), the ideal number of clusters seems to be two. The WSS method (middle) does not show any apparent elbow bend but suggests that the number of clusters should be between three and five. However, the silhouette method (right) suggests that five clusters would be the most suitable choice.

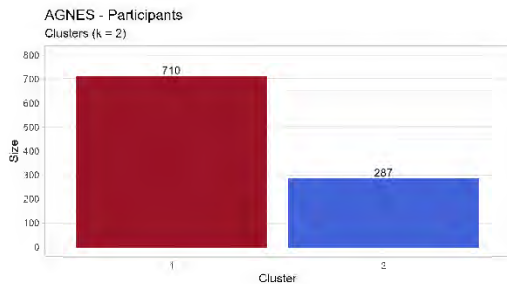


Figure 59: AGNES (k=2) size of participant clusters

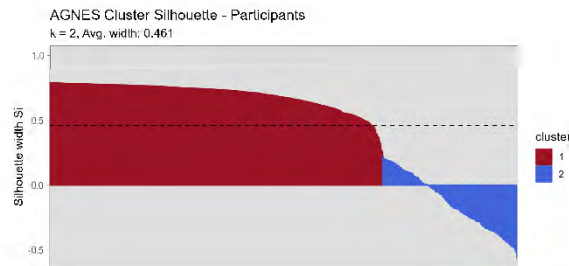


Figure 60: AGNES (k=2) cluster silhouette of participants

The cluster distribution, in Figure 59, is an improvement from K-means (k=2)(Figure 48) and is very similar to PAM (k=2)(Figure 56). The silhouette calculation in Figure 60 shows that some observations could be included in the other cluster. It also has a low silhouette average width similar to PAM's cluster choice (Figure 57). The full dendrogram is displayed in Figure 61 below.

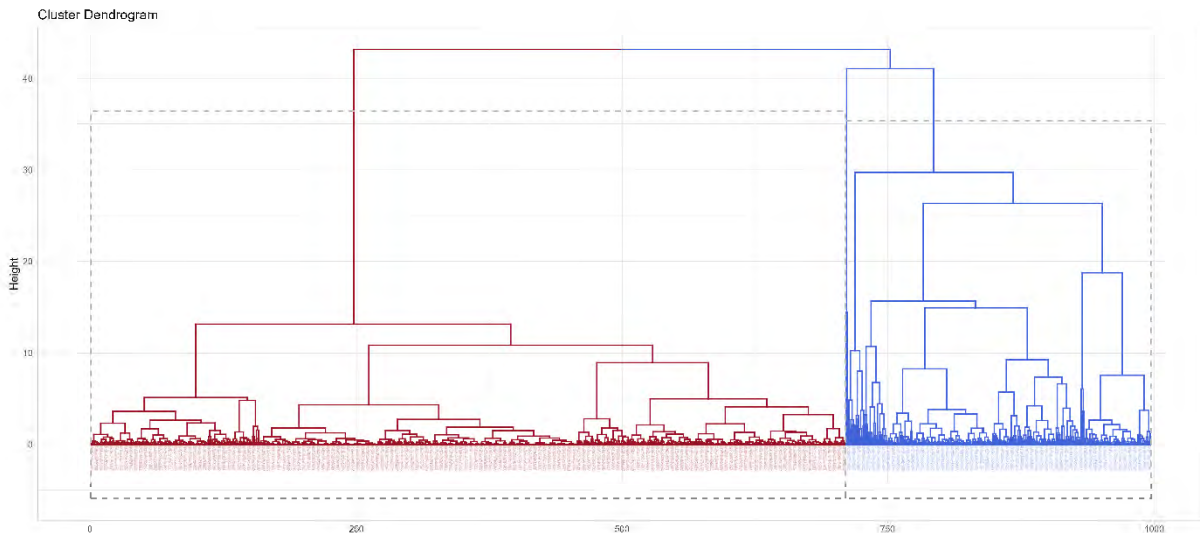


Figure 61: AGNES (k=2) Dendrogram

The mean of the roles (bar graph – Figure 62) is also very similar to what was seen in K-means and PAM, where the clusters are divided into sites with a large number of participants (cluster “2”) and a lower number of participants (cluster “1”). The conclusion that the cluster distinction is mainly based on the number of participants (mostly the number of students enrolled) is supported by Table 42. The table presents the minimum, mean, maximum, and standard deviation for each role group.

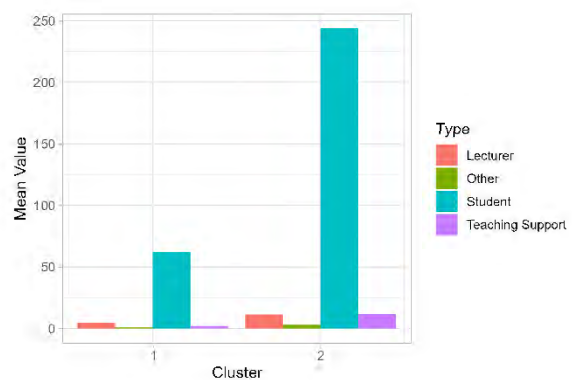


Figure 62: AGNES (k=2) mean of roles

Cluster	Lecturer	Student	Teaching Support	Other
1	1 : 4.606 : 15 (2.772)	4 : 62.054 : 353 (58.922)	0 : 2.052 : 9 (2.144)	0 : 0.844 : 16 (1.463)
2	1 : 11.411 : 77 (9.237)	10 : 243.732 : 1423 (235.432)	0 : 11.822 : 52 (9.417)	0 : 3.077 : 195 (13.602)

Table 42: AGNES (k=2) min:mean:max (standard deviation) of participants in summarised roles

4.2.2.3.2 Divisive Analysis Hierarchical Clustering (DIANA)

After running the divisive algorithm, the dendrogram is analysed using the gap statistics, total within sum of squares (WSS), and silhouette width comparison methods to determine the optimal number of clusters for cutting the tree. The results of the calculations (Figure 63) indicate that the gap statistics recommend a single cluster, the WSS suggests either three or five, and the silhouette method recommends two.

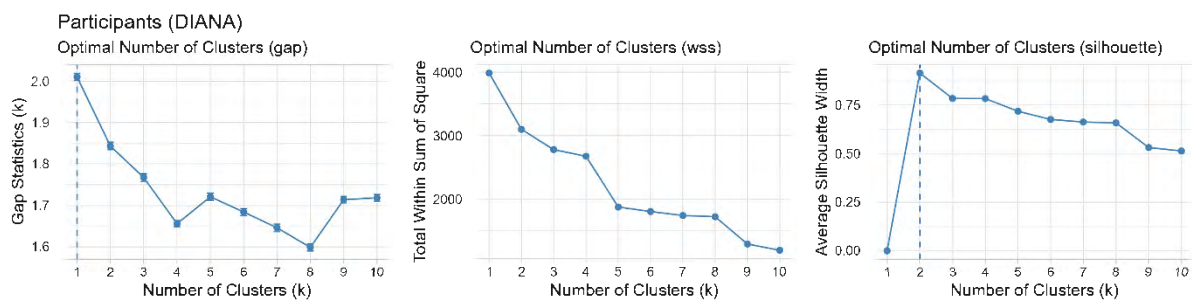


Figure 63: DIANA – Determine the optimal number of clusters using (gap, wss, silhouette) for participants

The resulting cluster sizes for k=2 (995:2), k=3 (990:5:2), and k=5 (961:29:5:1:1) indicate that the DIANA method is not the most optimal for the participant's data set. This is confirmed in the silhouette plots, which show the large cluster dominating the graph and the other clusters not near the average silhouette value (Cluster sizes and Silhouette graphs are shown in Appendix – Participant Data Set - DIANA).

4.2.2.4 Comparing Clustering Algorithms (Introducing clValid)

Selecting the most suitable clustering algorithm from the four popular algorithms is a challenging task, as is identifying the optimal number of clusters. The clValid R package (Brock et al., 2008) can assist in this process. This package allows for the simultaneous comparison of multiple clustering algorithms within a single function call, aiding the identification of the best clustering approach and the optimal number of clusters (or validation thereof).

In Section 2 of their paper, (Brock et al., 2008) describes the three validation methods (“internal”, “stability”, and “biological”) that can be used to compare clustering algorithms (page 7). Internal validation measures evaluate the quality of a clustering result based on intrinsic information in the data set and the clustering partition. Stability measures are a type of internal measure that assesses the consistency of a clustering result by comparing it with the clusters obtained after each column is removed, one at a time. Biological validation evaluates the clustering algorithm to produce biologically meaningful clusters (Brock et al., 2008); this evaluation method does not fit into the scope of this study.

The internal validation method has three metrics to evaluate the clustering algorithm on, namely “Connectivity”, “Dunn”, and “Silhouette”.

Connectivity is the combined score of compactness and separation, as described in (Brock et al., 2008). Compactness (cluster cohesion) is a measure of how close the observations are within each

cluster. A lower within-cluster variation indicates good compactness, i.e., good clustering (Kassambara, 2023c). Separation is the measure of how well the clusters are separated from one another, and this measure includes the distance between cluster centres and the minimum between observations in different clusters (Kassambara, 2023c). These two measurements are inversely related – compactness increases with the number of clusters, but separation decreases (Brock et al., 2008). This value should be as low as possible for the least number of clusters, indicating the optimal number.

The Dunn index is the ratio of the smallest distance between observations not in the same cluster to the largest intra-cluster distance (Brock et al., 2008). The Dunn index has a value between zero and ∞ and should be maximised for the lowest number of clusters.

The Silhouette width calculation indicates how well an observation is clustered. The value used is the average width for all observations (Brock et al., 2008). This is similar to the silhouette graphs used in the previous analysis steps. Silhouette values range from -1 to 1. Closer to 1 indicates good clustering.

The comparison method (clValid) is run for the selected clustering algorithms K-means, PAM, Hierarchical (AGNES), and DIANA. The detailed calculations for each measure and method for cluster sizes from two to ten are tabled below.

Measure	Method	2	3	4	5	6	7	8	9	10
Connectivity	diana	4.858	43.992	44.992	56.036	60.889	67.026	68.567	130.792	143.911
Connectivity	hierarchical	4.858	11.190	12.190	30.198	35.660	68.227	70.377	77.014	146.945
Connectivity	kmeans	4.858	70.579	71.579	83.191	139.911	164.217	193.359	198.611	193.313
Connectivity	pam	130.573	208.627	263.410	285.998	339.480	418.280	378.543	429.513	441.973
Dunn	diana	0.638	0.023	0.030	0.031	0.036	0.038	0.045	0.026	0.028
Dunn	hierarchical	0.638	0.145	0.186	0.096	0.101	0.046	0.049	0.054	0.018
Dunn	kmeans	0.638	0.016	0.016	0.017	0.013	0.011	0.008	0.002	0.010
Dunn	pam	0.004	0.001	0.001	0.001	0.000	0.000	0.000	0.001	0.001
Silhouette	diana	0.909	0.697	0.697	0.683	0.684	0.652	0.645	0.518	0.487
Silhouette	hierarchical	0.909	0.749	0.748	0.724	0.714	0.561	0.559	0.479	0.224
Silhouette	kmeans	0.909	0.610	0.609	0.608	0.439	0.439	0.416	0.406	0.405
Silhouette	pam	0.471	0.292	0.207	0.201	0.207	0.187	0.214	0.189	0.153

Table 43: Participant cluster comparison (Internal) – Table

Figure 64 (on the next page) contains the measures described in the table above; it makes it easier to visualise the change in values for each clustering method. The connectivity graph at the top indicates that two clusters for DIANA, AGNES, and K-means is the optimal choice (connectivity should be as low as possible for the least number of clusters). For the Dunn index (middle), the value should be maximised for the least number of clusters, in this case, two for K-means, AGNES, and DIANA. Similarly, the Silhouette width should be close to 1 with the lowest cluster count, which is two clusters for the same selection. Reviewing the actual values for each of the internal evaluation methods (in the table below) shows that for two clusters, the values for K-means, AGNES, and DIANA are identical, which indicates that 2 clusters are likely the optimal choice for any of these methods.

Participants - Internal Validation

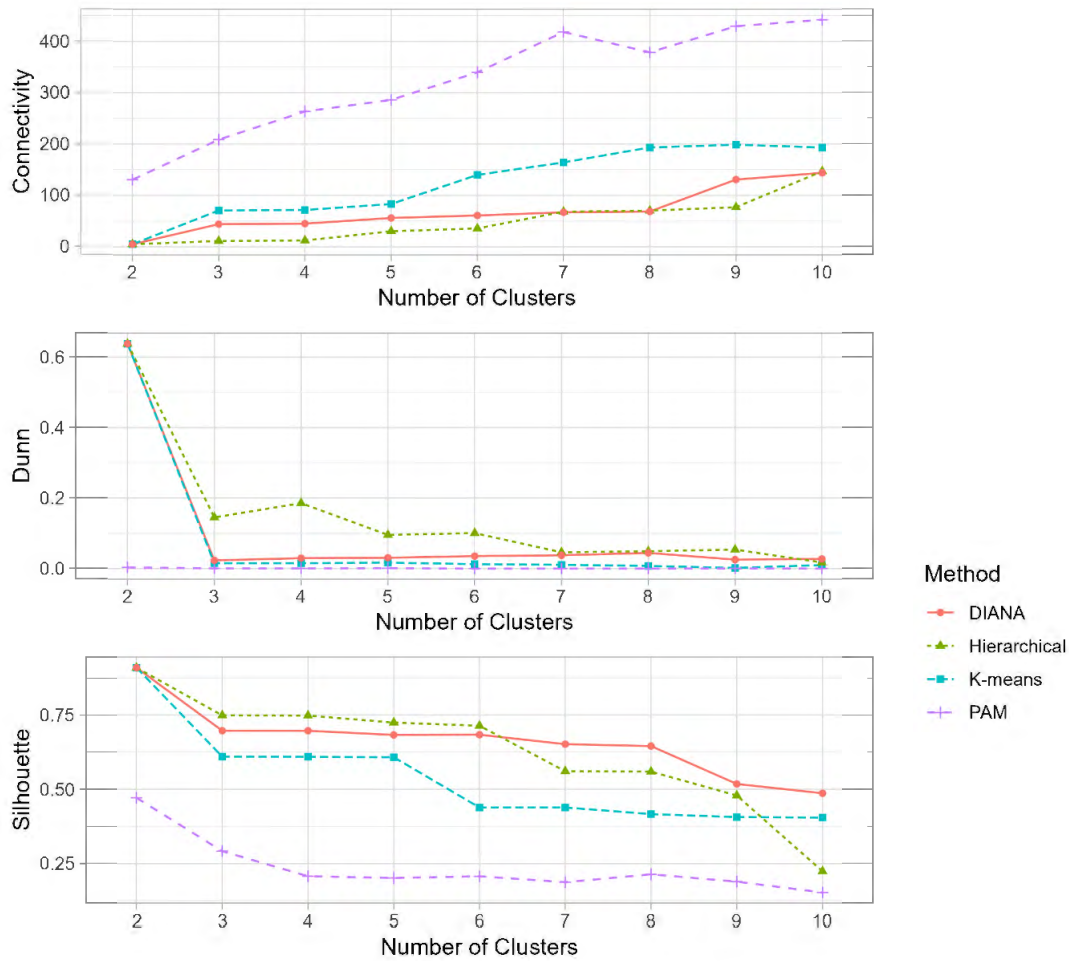


Figure 64: Participant cluster comparison (Internal) – Graphs

The internal validation method gives an accurate measure of optimal clusters comparable to the gap statistics, the total within sum of squares, and the silhouette method used in the sections above. The rest of the analysis phase data sets will first be analysed with the cValid internal validation method to indicate which clustering algorithms fit the data set the best.

The stability validation measure evaluates the consistency of the clustering results by using the full data set compared to the same data set with one variable (column) removed (Brock et al., 2008; Kassambara, 2023c). The measure includes the average proportion of non-overlap (APN), the average distance (AD), the average distance between means (AD), and the figure of merit (FOM). The average is taken in all measures based on the full data set and then again with each subsequent variable (column) removed. All measures should be minimised (Brock et al., 2008).

- The APN measures the average proportion of observations not placed in the same cluster under both cases (full data set and removal of one variable). Values approaching 0 indicate highly consistent clustering results.
- The AD measures the average distance between observations in the same cluster under both cases. This has a value between 0 and ∞ (smaller values are preferred).

- The ADM measures the average distance between cluster centres for observations in the same cluster under both cases. This has a value between 0 and ∞ (smaller values are preferred).
- The FOM measures the average intra-cluster variance of the deleted variables (columns), where the clustering is based on the remaining (undeleted) variables. This has a value between 0 and ∞ , with smaller values showing better performance.

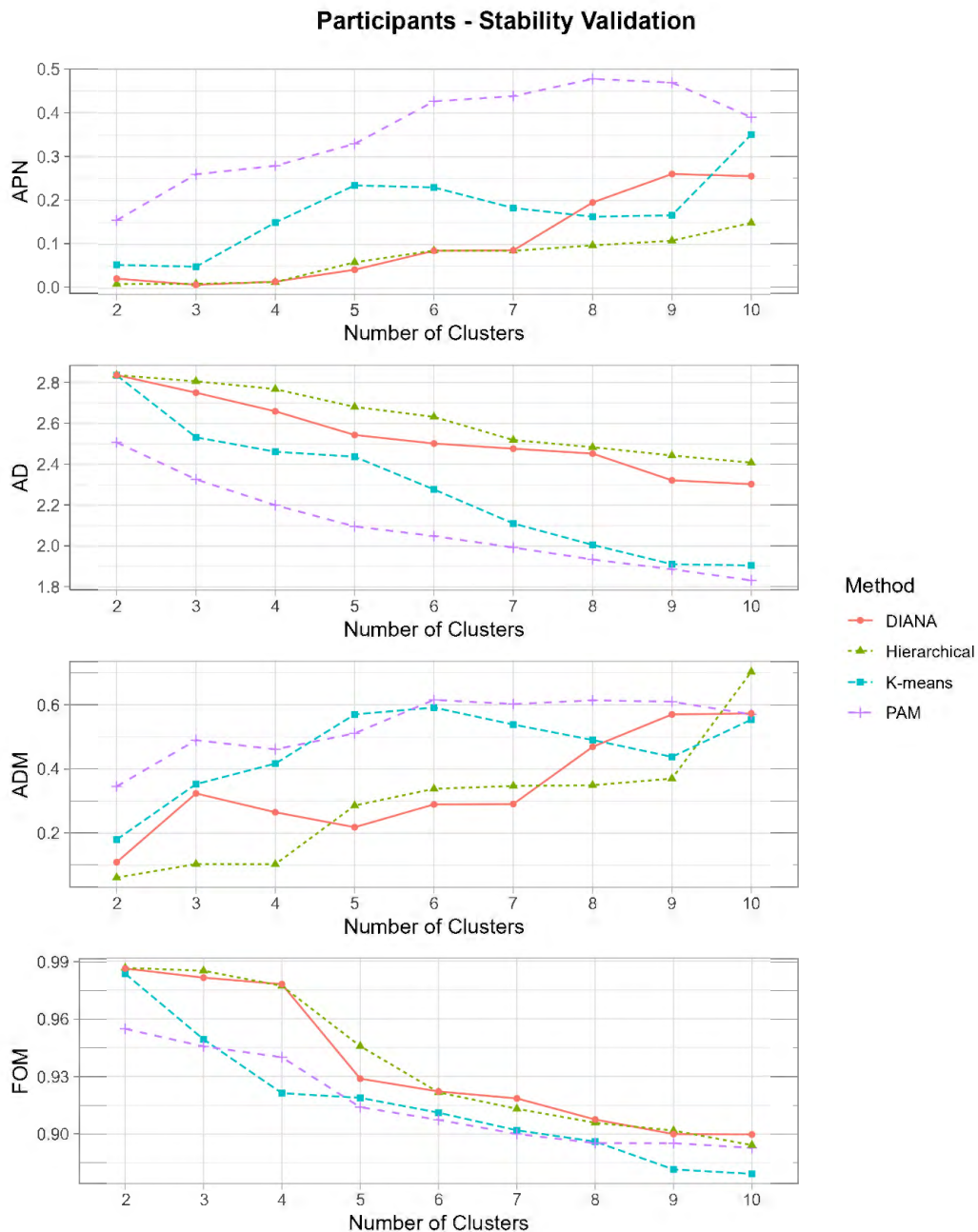


Figure 65: Participant cluster comparison (Stability) – Graph

The following table (Table 44) summarises the selection of the lowest value for each stability validation measure and the lowest number of clusters identified using the participant data set. Unfortunately, this suggests that a different clustering method performs better for each measure. As a result, for the remainder of the analysis, the internal validation method will be utilised as it provides a more consistent recommendation on algorithm and number of clusters.

	Score	Method	Clusters
APN	0.007774	diana	3
AD	1.832058	pam	10
ADM	0.061567	hierarchical	2
FOM	0.879290	kmeans	10

Table 44: Participant (clValid) Stability

4.2.2.5 Comparing Clusters to Event Data

While it is not the primary objective of this study, analysing the number of participants in a course site in relation to the number of events can provide valuable insights into the data selection process and its validity. Before comparing the clusters identified in the preceding sections, the level of student interaction in all course sites with the total number of students enrolled in those sites will be compared.

The number of students positively correlates to the total number of events for course sites (Pearson = 0.602, Spearman = 0.844). Figure 66 (below) shows the total number of students against the total number of events. Most sites have a low number of students with a low number of interactions for those students. The outliers “STA1000S,2019” (904) (1,227 students, 5 lecturers, and 7 tools) and “BUS1036F,2019” (649) (942 students, 2 lecturers, and 27 tools) have been discussed in the data extraction section and re-occur here due to the enormous number of events generated by the tools for the students in the sites.

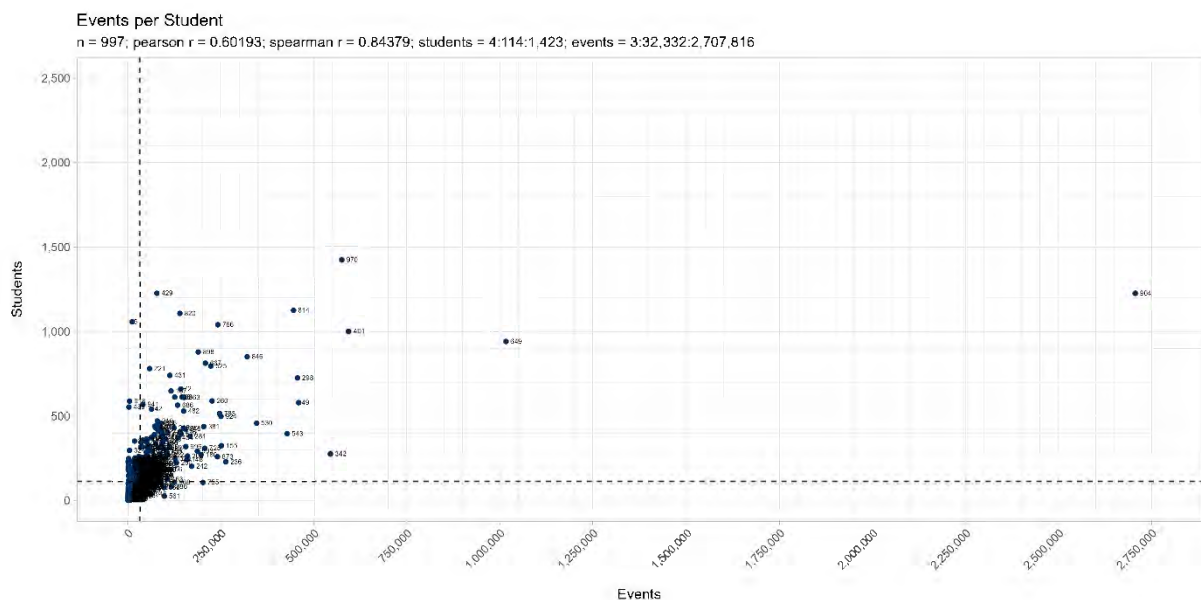


Figure 66: Students to the total event in course sites

Figure 67 (below) maps students to the average event per student in a course site. The outlier sites with a high number of students, “ECO1011S 2019” (970) (1,423 students, 16 lecturers, and 23 tools) and “STA1000S,2019” (904) (see above), have very similar tool structures, with a high percentage of usage by students. The other atypical site, “STA2007F,2019” (561), has 28 students enrolled and a very high number of events associated with assessment and content (the site uses resources and lessons extensively).

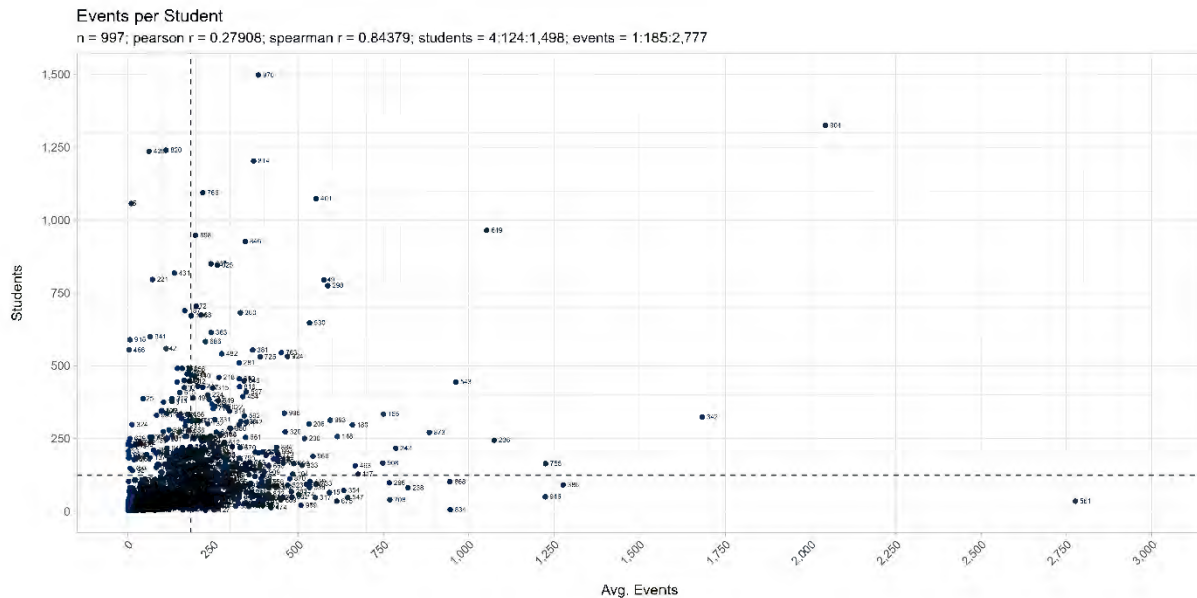


Figure 67: Students to the average event per student in course sites

After conducting the clustering analysis, it was recommended to use 2 clusters for the K-means, AGNES, and DIANA algorithms based on the results of the internal validation method. The resulting cluster sizes for the three chosen algorithms are reproduced in Figure 68 (below). The DIANA algorithm's final cluster size of 995:2 (below – right) is not a reasonable result and as such, is excluded from further analysis.

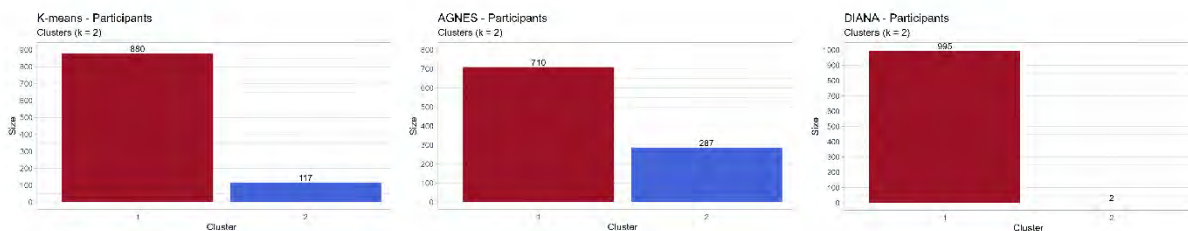


Figure 68: K-means, AGNES, DIANA cluster sizes for participants (k=2)

Table 45 and Table 46 display the sites closest to the cluster centres for K-means and AGNES, respectively.

K-means:

Site Title	Cluster	Distance to Center	Lecturer	Student	Teaching Support	Other
Orchestration 1, 2019	1	2 3713	1	4	1	0
Composition 1, 2019	1	2 3713	1	4	1	0
Molecular Medicine 2019	2	52 1197	51	17	1	0

Table 45: Cluster min:mean:max (sd) for K-means (k=2) for participants

Hierarchical Clustering (AGNES):

Site Title	Cluster	Lecturer	Student	Teaching Support	Other
STA2007H,2019	1	4	4	6	0
ACC1106F 2019	2	9	258	15	1

Table 46: Cluster min:mean:max (sd) for AGNES (k=2) for participants

In the results from both algorithms, it is evident that the distinction between the clusters are based on the number of “Lecturer” and “Student” roles in each site. Cluster 1 contain the sites with a low number of lecturers and students and Cluster 2 contains more users enrolled as lecturer and students.

4.2.3 Tools Used in Course Sites

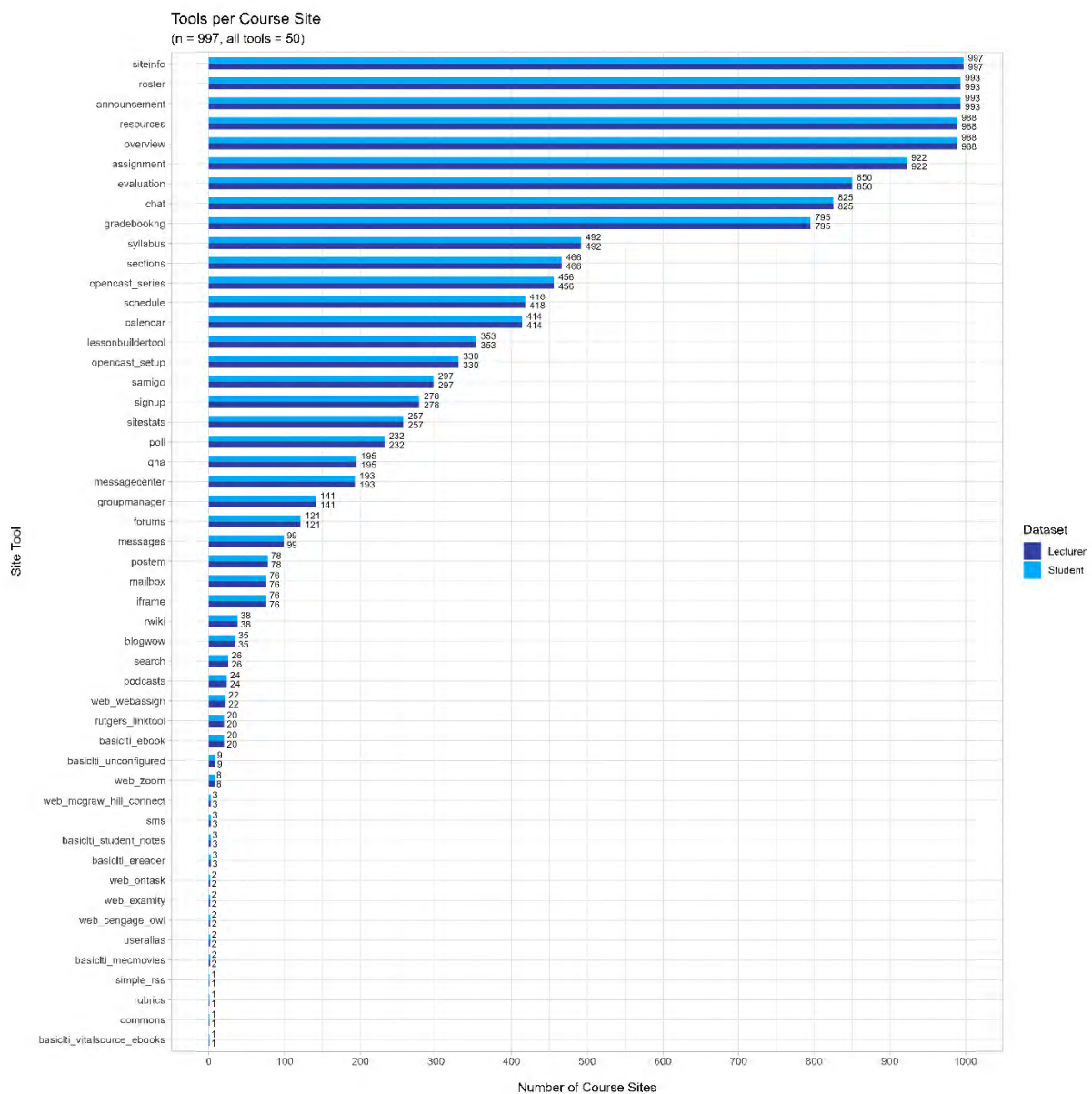


Figure 69: Tools visible to Lecturers and Students in the filtered course sites.

Examining the number of all the tools (visible to lecturers and students - Figure 69) in the course sites, the majority of sites still use the default assigned tools - Announcements, Roster, Resources, Chat, etc. The site information (`siteinfo`) tool is right at the top as it is a required tool for all course sites; it's used to administer all the tools and participants on the site. It is plausible that many lecturers would include assignments in their sites in conjunction with the grade book. 326 Course sites did not utilise the feature "Welcome to Lecture Recording" (`opencast_setup:LTI`) but kept the tool link, 456 transformed their link into "Lecture Recording" (LTI), and 211 just removed the tool. At the bottom of the graph, the tools used less frequently were the more bespoke LTI and web integrations.

Mapping the student visible tools in a site to the enrolled students, the same pattern emerges. Most sites have the default communication, assessment, management, and content tools. Note how sites with close to the mean number of students (or smaller) tend to implement more non-standard tools, possibly because the tools are new or because the class size allows for a more creative pedagogy. There are exceptions; "Student Notes" (`basiclti_student_notes`), which is used in the Department of Mathematics and Applied Mathematics – to store communication between lecturers (visible at the bottom of Figure 69 and Figure 70).

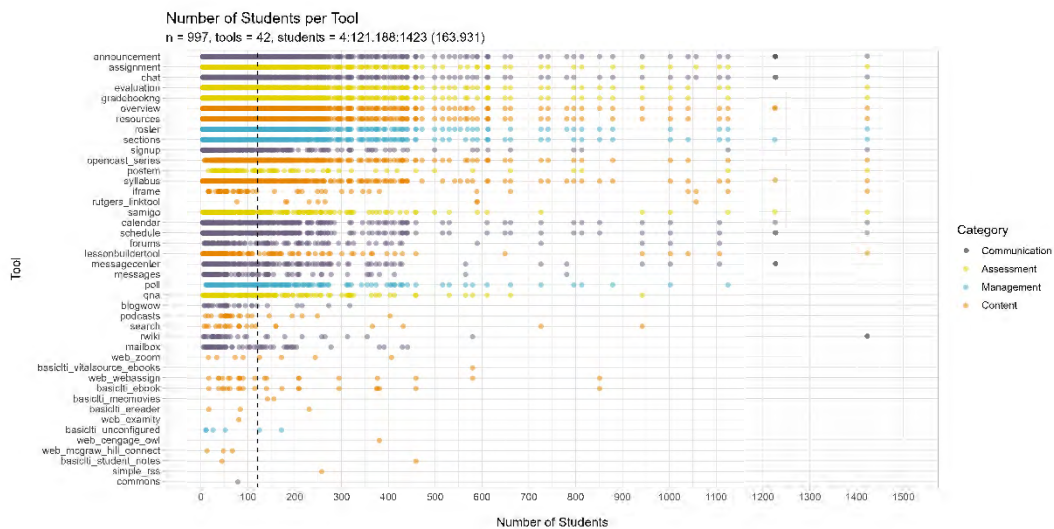


Figure 70: Number of students to the tool choices grouped by category

Figure 71 represents all the events per tool per category; this shows that the content tool resources are by far the most used tool, followed by the main assessment tools (`samigo`, `gradebookng`, `assignment`). Announcements are fifth on the list, followed by the content tool for lecture recording (`opencast_series`).

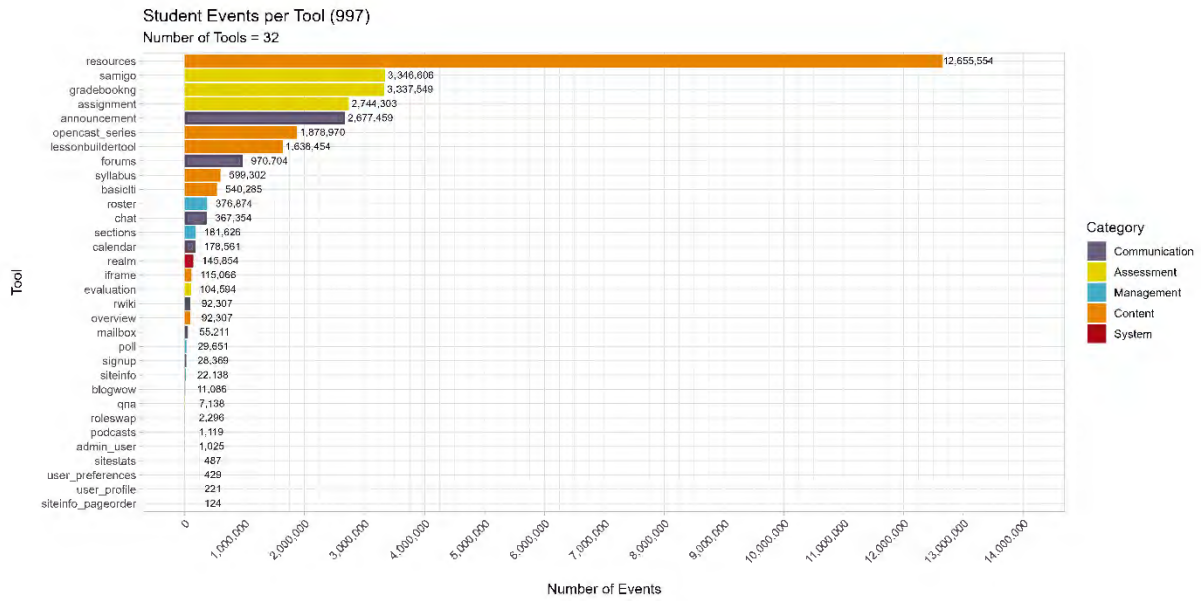


Figure 71: Number of events per tool grouped by tool category

Category	Event Type	Total Events	Total Students	Avg. Event per Student	No. Tools
Content	32	17,521,057	18,113	967.319	8
Assessment	97	9,540,190	17,799	535.996	5
Communication	62	4,381,051	17,940	244.206	8
Management	31	610,776	17,060	35.802	5
System	17	149,949	13,540	11.075	6

Table 47: Event, student, and tool details per category

Table 47 shows the tool category, the number of distinct event types followed by the total number of events generated in that category by students, the total number of unique students that generated these events and the average number of events per student. Lastly, it depicts the total number of tools included in that category out of the 32 tools visible to students tracked in the events table for course sites. It's evident from the table that the content tools generate the majority of events for course sites (more than the other categories combined) even though the types of events tracked are on the lower end of the scale.

In Figure 71 (previous page) the content tool occurs five times in the top 10, indicating a lot of events generated for the category, but only occurs once in the top 10 tools selected by lecturers (resources - Figure 69).

Events	Students	Student Tools
3 : 32,332.352 : 2,707,816	4 : 114.463 : 1,423	3 : 12 : 32

Table 48: min:mean:max for events, students and tools visible to students in course sites

The average course site generates around 32 thousand events, contains about 12 tools, and, on average, has 114 students enrolled. Figure 72 shows the correlation between the number of tools in a course site and the enrolled students. For the 997 course sites and the 32 tools visible to students shown in the graph, the Spearman correlation coefficient indicates a moderately weak positive relationship (0.209), indicating that there is a modest correlation. The Pearson correlation coefficient

of 0.207 confirms this relatively weak correlation between the number of students and the number of tools selected.

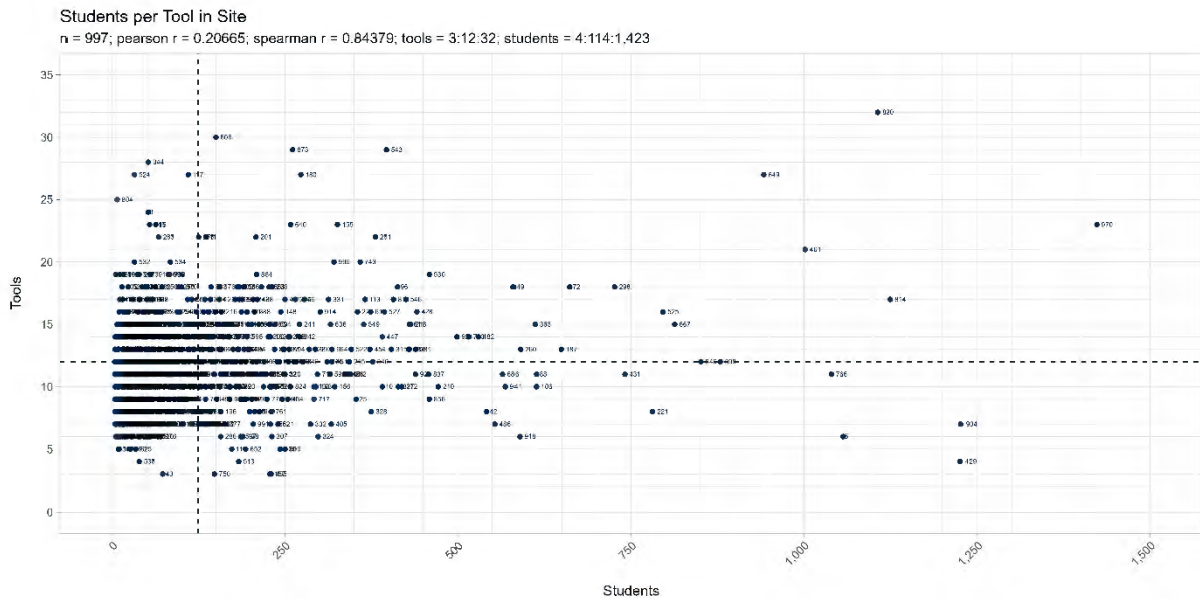


Figure 72: Students' correlation to tools

Still, it's more likely that the tool selection depends on the default set of tools added to the site on creation or the lecturer's pedagogical approach, available teaching support, system constraints, and/or time constraints.

Comparing the number of tools available for two main roles, namely "Lecturer" (All Tools) and "Student" (Tools Visible to Students), produces the following results. The histogram in Figure 73 (below) shows the data sets for both roles. The distribution of the number of tools per course site is very similar, with minor variations due to restricted access to certain tools such as grading and assessment tools, course management, and system administrator tools.

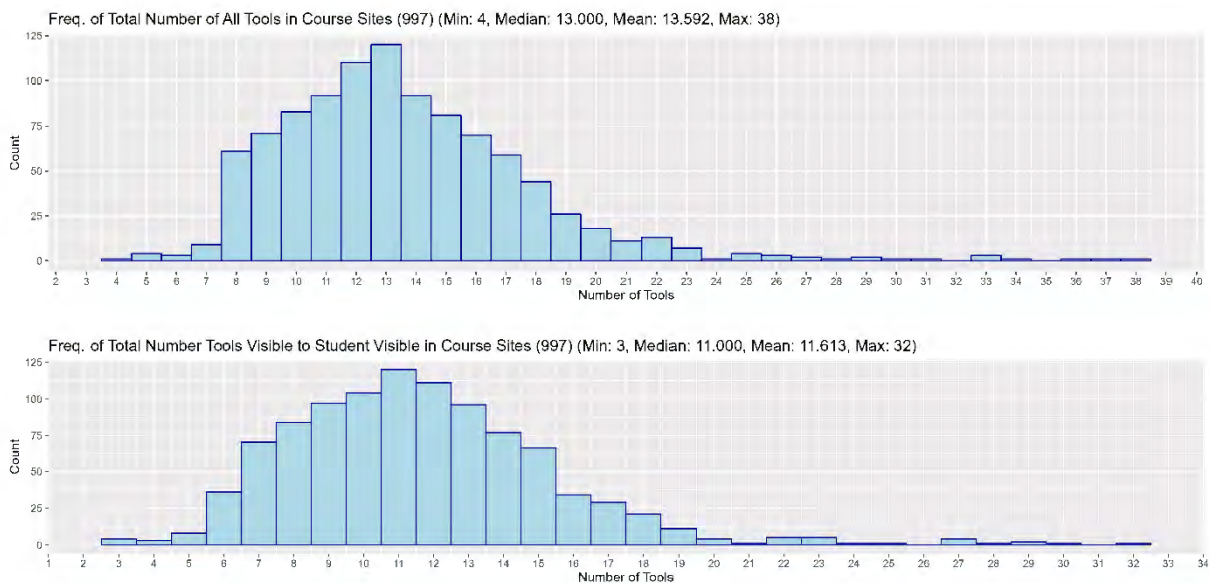


Figure 73: Distribution of the tools in course sites (all tools or only visible to students).

Looking at the Hopkins statistic of the data sets “All Tools” and “Tools Visible to Students”, both values are close to 1, indicating low spatial randomness and thus a high clusterability.

All Tools (49 Tools)	Tools Visible to Students (42 Tools)
\$hopkins_stat [1] 0.94383	\$hopkins_stat [1] 0.94312

Table 49: Hopkins statistics for all tools and tools visible to students’ data set.

This study focuses on students interacting with tools chosen per site rather than tools visible to all roles. The "All Tools" dataset, while interesting, is not used in the rest of this analysis.

Below are two tables, Table 50 and Table 51, listing the tools available to students. The tables are categorised and ordered by the number of tools, starting with the lowest 10 and the highest 10. It is evident that students have limited access to course management (Management) and system administrator tools (System).

Title	Department	Faculty	Assessment	Communication	Content	Management	System	# Tools
EDN2521W,2019	EDN	HUM	1	1	1	0	0	3
Assess HUB1006F,2019	HUB	FHS	1	1	0	1	0	3
Assess IBS1007S, 2019	IBS	FHS	1	0	1	1	0	3
INF1102S_Makeup Test	INF	COM	1	0	1	1	0	3
Tests STA1000S,2019	STA	SCI	1	1	1	1	0	4
DOL4000H,2019	DOL	LAW	1	1	1	1	0	4
INF3003W, Final Exam 2019	INF	COM	1	0	2	1	0	4
EEE4122C,2019	EEE	EBE	1	1	2	1	0	5
FIN3025W,2019	FIN	HUM	1	1	2	1	0	5
MAM2083S,2019	MAM	SCI	0	1	4	0	0	5

Table 50: Course site sorted by total number of tools visible to students (Total Tools - lowest 10).

Title	Department	Faculty	Assessment	Communication	Content	Management	System	# Tools
Discover Commerce 2019	DOC	COM	2	5	22	3	0	32
INF1102S,2019	INF	COM	5	6	15	4	0	30
INF1002F - 2019	INF	COM	5	5	17	2	0	29
INF1002S - 2019	INF	COM	4	6	17	2	0	29
AHS1035F,2019	AHS	FHS	6	10	7	5	0	28
INF1102F,2019	INF	COM	5	4	14	4	0	27
INF2009F,2019	INF	COM	6	6	12	3	0	27
AHS4008/9H,2019	AHS	FHS	6	6	11	4	0	27
BUS1036F,2019	BUS	COM	5	5	16	1	0	27
HSE 1014S and HSE 10 15F	HSE	FHS	4	8	11	2	0	25

Table 51: Course site sorted by total number of tools visible to students (Total Tools - top 10)

The data set “Tools Visible to Students” contains 42 tools. In an effort to reduce the number of variables (dimensions) needed for cluster analysis and increase the interpretability of the data but still maintain the variance with minimal information loss (Keita, 2023). A Principal Component Analysis (PCA), a popular unsupervised learning technique, was applied to this particular dataset. The results showed a slight reduction in variables from 42 to 32 (a decrease of 23.809%), this did not

significantly impact the clustering or interpretability of the data. The results are included in Appendix – Tools Visible to Students - PCA.

4.2.3.1 Tools Visible to Students - Clustering

As explained earlier, the cIValid package and the internal measures can assist in selecting the number of clusters and the clustering algorithm.

Measure	Method	2	3	4	5	6	7	8	9	10
Connectivity	diana	2.929	5.858	8.787	11.716	14.645	18.503	22.885	26.743	31.030
Connectivity	hierarchical	2.929	5.858	8.787	14.645	14.645	18.503	30.935	30.935	35.317
Connectivity	kmeans	2.929	8.787	11.716	11.716	14.645	18.503	30.935	30.935	35.317
Connectivity	pam	457.567	491.785	638.430	684.185	638.393	633.397	612.442	680.446	722.412
Dunn	diana	0.668	0.720	0.737	0.740	0.793	0.706	0.491	0.508	0.523
Dunn	hierarchical	0.668	0.720	0.737	0.742	0.793	0.706	0.491	0.503	0.512
Dunn	kmeans	0.609	0.619	0.654	0.740	0.793	0.706	0.491	0.503	0.512
Dunn	pam	0.029	0.013	0.013	0.013	0.014	0.014	0.014	0.032	0.032
Silhouette	diana	0.793	0.788	0.780	0.771	0.763	0.724	0.627	0.630	0.630
Silhouette	hierarchical	0.793	0.788	0.780	0.767	0.763	0.724	0.632	0.619	0.619
Silhouette	kmeans	0.770	0.770	0.771	0.771	0.763	0.724	0.632	0.619	0.619
Silhouette	pam	0.070	0.063	0.044	0.048	0.059	0.063	0.069	0.065	0.051

Table 52: Tools visible to students' cluster comparison (Internal) – Table

The clustering algorithms K-means, AGNES, and DIANA produce the lowest connectivity value when selecting 2 clusters. On the other hand, the best Dunn index value is obtained when these three algorithms are used with six clusters. Furthermore, the silhouette width measure shows that AGNES and DIANA performed the best for the two clusters.

The values in the table above are visually plotted in the graph (Figure 74) on the next page - in all the measures, the K-medoids (PAM) clustering algorithm performed by far the worst. The graphs highlight the conclusions expressed in the table that the other three clustering algorithms perform almost the same (at least for these measures). In this case, the K-means algorithm would be the most suited for this data set as it is far less computationally intensive than the hierarchical clustering methods (AGNES and DIANA).

Tools Visible to Students - Internal Validation

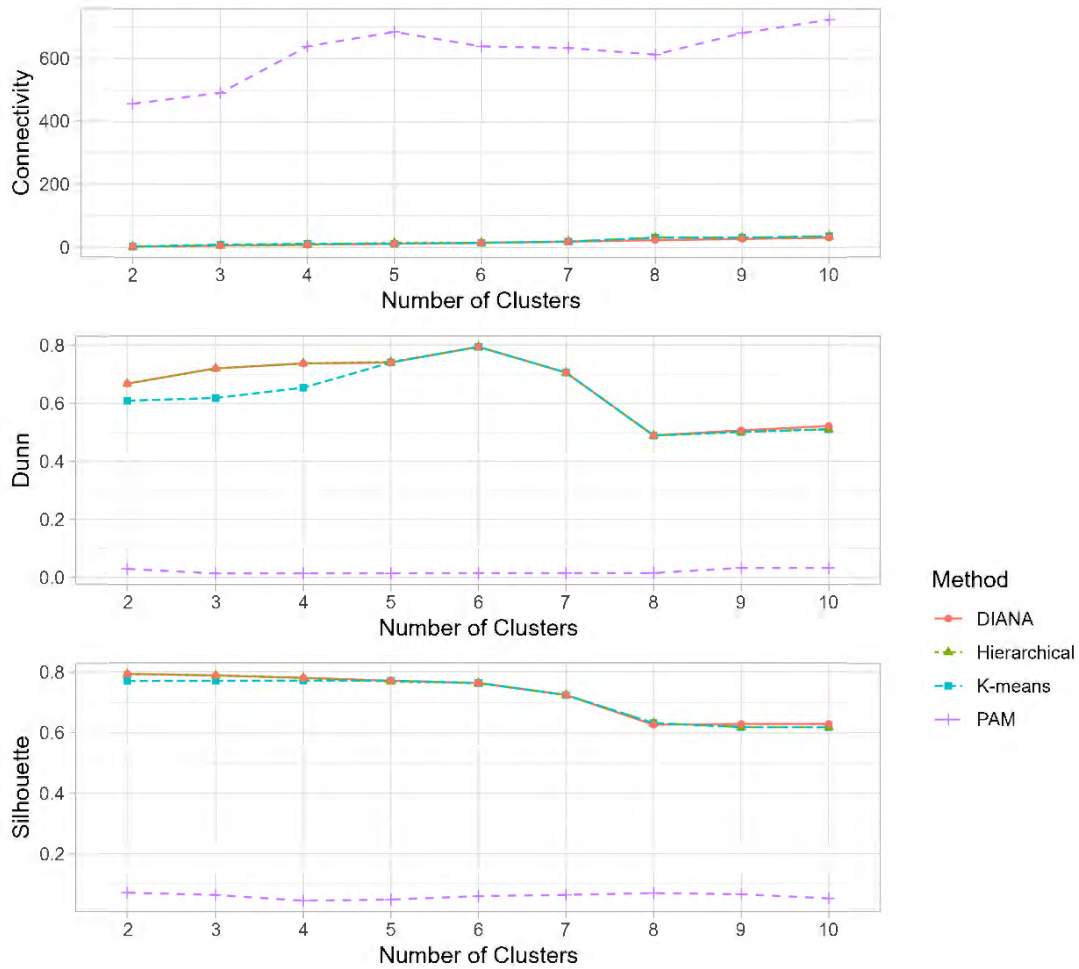


Figure 74: Tools visible to students' cluster comparison (Internal) – Graphs

The recommended cluster algorithms and sizes were used to produce the following cluster size and silhouette graphs for all the tools visible to students in the selected course sites. Unfortunately, the DIANA algorithm produced very poor clustering results (below) and will be excluded from the analysis in this section.

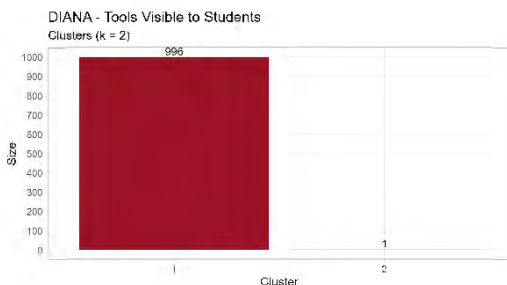


Figure 75: DIANA (k=2) cluster size for tools visible to students

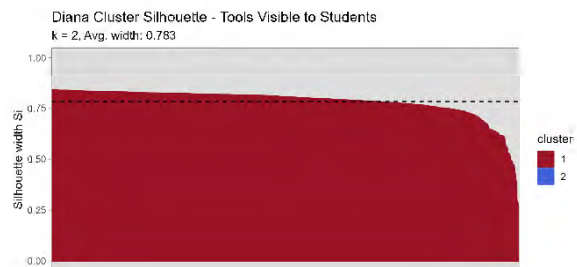


Figure 76: DIANA (k=3) cluster silhouette for tools visible to students

The AGNES and K-means clustering produces reasonable cluster breakdowns for k=2, but the silhouette graph indicates misallocated observations in the second cluster.

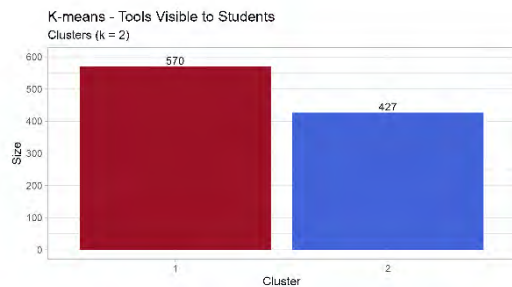


Figure 77: K-means (k=2) cluster size for tools visible to students

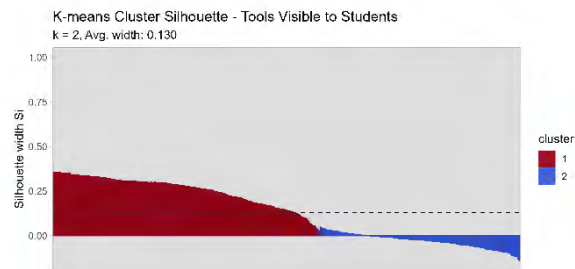


Figure 78: K-means (k=3) cluster silhouette for tools visible to students

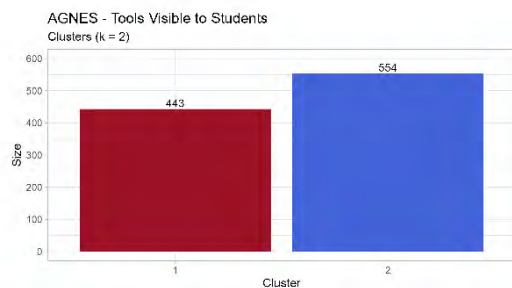


Figure 79: AGNES (k=2) cluster size for tools visible to students

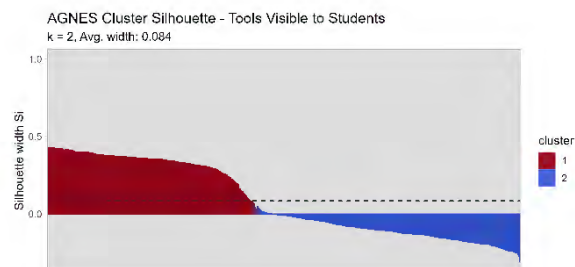


Figure 80: AGNES (k=3) cluster silhouette for tools visible to students

Looking at the breakdown of tools based on the category of each algorithm, it seems that there is not much difference between the two methods.

K-means:

Cluster	Assessment	Communication	Content	Management
1	0 : 2.774 : 6 (1.133)	0 : 2.326 : 6 (0.997)	0 : 3.023 : 8 (1.036)	0 : 1.488 : 4 (0.647)
2	0 : 3.543 : 7 (1.032)	3 : 5.063 : 11 (1.305)	1 : 3.817 : 22 (2.319)	0 : 1.862 : 4 (0.717)

Hierarchical Clustering (AGNES):

Cluster	Assessment	Communication	Content	Management
1	0 : 2.582 : 6 (1.107)	0 : 2.097 : 5 (0.765)	0 : 2.828 : 15 (1.058)	0 : 1.429 : 3 (0.628)
2	0 : 3.520 : 7 (1.016)	1 : 4.619 : 11 (1.530)	2 : 3.791 : 22 (2.054)	0 : 1.823 : 4 (0.710)

Table 53: Cluster min:mean:max (sd) for K-means and AGNES (k=2) for tools visible to students

Table 53: Cluster min:mean:max (sd) for K-means and AGNES (k=2) for tools visible to students Table 53 (above) contains a table describing each of the clusters obtained from using K-Means and AGNES where k=2 and the categories of tools used in each cluster. The column for each category shows the minimum, mean, and maximum and the standard deviation of the number of tools in each category. The two clusters in the two clustering algorithms are very similar across all the categories, with the same minimum and maximum tools and close to similar mean and standard deviation.

Looking at the tables for the closest observations to the centre of the clusters (Table 54), the K-means results aren't as promising; the only categories with differences are communication and content.

Site Title	Cluster	Distance to Centre				
		Assessment	Communication	Content	Management	
CML1001L,2019; GEO1006S,2019; Greek IB 2019; AHS1033F,2019; SLL3131F,2019; CML2001L,2019; CML4501F,2019; AST3002F,2019; SLL2131F,2019; RDL1007F,2019; SLL3002H,2019; SLL1042F,2019; SLL1048H,2019; MEC4105F,2019	1	1.8299	3	2	2	1
HUB1023S,2019	2	2.2401	3	5	4	1

Table 54: K-means (k=2) sites close to the centre of clusters for tools visible to students

Table 55 shows the results for the AGNES clustering algorithm, where the difference between the two representative sites is more well-defined. Cluster 1 is represented by “Tests STA1000S, 2019”, a site that was described in the previous section because of its low number of tools and the high number of students enrolled.

Site Title	Cluster	Assessment	Communication	Content	Management
Tests STA1000S,2019	1	1	1	1	1
PBL4505S,2019	2	2	5	3	2

Table 55: AGNES (k=2) sites close to the centre of clusters using tools visible to students

In the next section, these cluster results will be used and compared with the student interactions for the representative sites for each of the clustering algorithms.

4.2.3.2 Tools Visible to Students - Comparing Clusters to Event Data

Comparing the clusters of tools selected to event data is the main purpose of this study, trying to determine a specific usage pattern for tool selection that leads to students' interaction with those tools.

It's important to understand the type of events that are recorded for student interactions as this is a system restriction due to the LMS architecture and development decisions. Each tool creates an entry for a specific interaction that the software developer decided was a meaningful event for that tool, this leads to a varying number of events per tool and category. Event information is not available for external tools, so although there are a number of LTI tools available to students, the interactions are stored outside of the LMS and are not available in this analysis.

Figure 81 (next page) displays the total number of events observed for students in the 2019 course sites (Event) and compares it to the number of possible events tracked by Sakai (All). The difference between tracked and possible events may be due to students' limited access to tool activities (e.g., creating vs. reading lesson content).

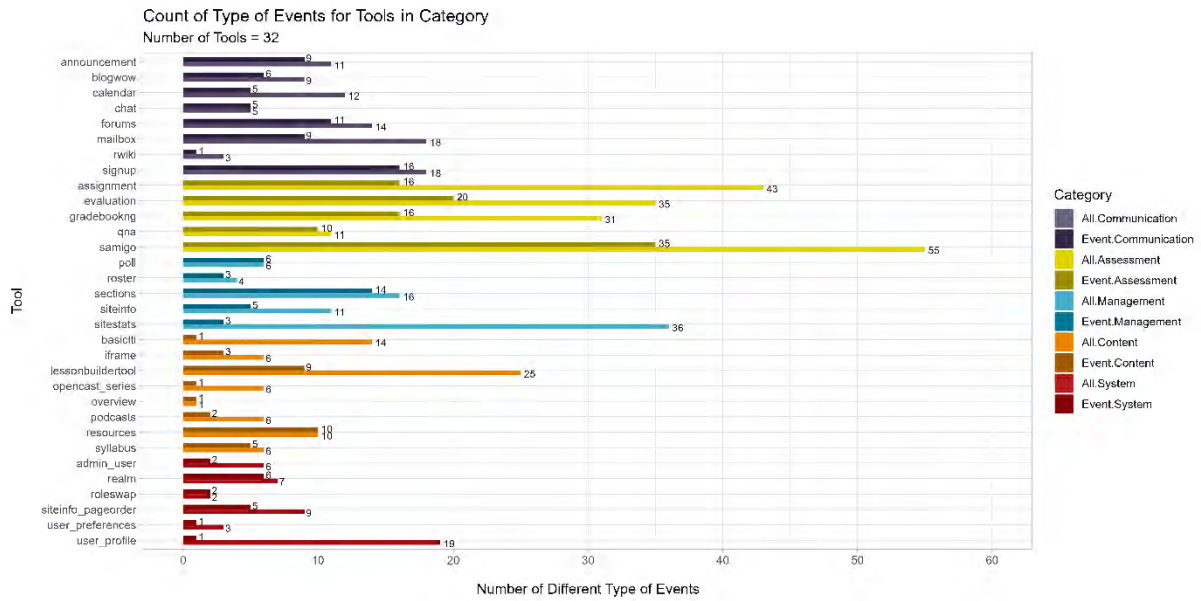


Figure 81: Count of event types in event data

Another factor that influences the number of events is the participants (students) in the course sites. This analysis focuses mostly on student interactions. Further study could examine the influence of lecturers or teaching support.

A low positive correlation exists between having more tools in a site and having more events (Pearson = 0.3727, Spearman = 0.467). This is not necessarily causation; the increased number of tools might be related to additional communication tools (causation) or the lecturer deciding that chunking content is more appropriate for the course (not causation, pedagogical choice).

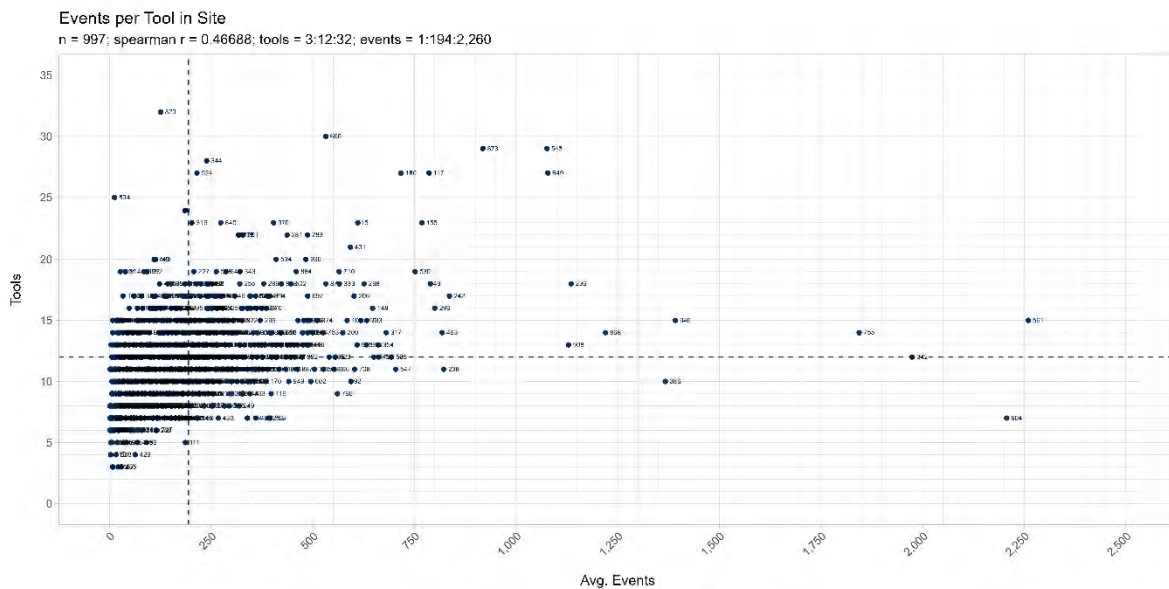


Figure 82: Number of tools on a site to total events

The four sites with a large number of events (right side) close to the mean number of tools are “STA2007F,2019” (561 - 28 students; 15 tools), “STA1000S,2019” (904 - 1,227 students; 7 tools), “STA1000F,2019” (342 – 276 students; 12 tools), and “CHE2005W,2019” (755 – 109 students; 14 tools). The site with the most tools is “Discover Commerce 2019” (820 – 1,107 students; 32 tools).

Returning to the clustering results of K-means ($k=2$) (Table 54) which had 2 clusters, 540 and 457 in size (Figure 77). The silhouette graph (Figure 78) shows the average silhouette weight is quite low and only crosses the first cluster. The sites identified to be closest to the centres have the same tool configurations for all the sites in Cluster 1 (14 sites) compared to a similarly structured site in Cluster 2. The sites in Cluster 1 have a diverse participant structure (Table 56) and across multiple departments, but they are all consistent in the number of tools that they contain (Table 57).

Title	Cluster	Total Tools	Lecturer	Student	Teaching Support
AHS1033F,2019	1	8	12	45	4
AST3002F,2019	1	8	2	15	1
CML1001L,2019	1	8	7	70	0
CML2001L,2019	1	8	5	57	0
CML4501F,2019	1	8	3	25	1
GEO1006S,2019	1	8	4	12	0
Greek IB 2019	1	8	4	5	0
MEC4105F,2019	1	8	3	7	1
RDL1007F,2019	1	8	3	22	0
SLL1042F,2019	1	8	3	13	0
SLL1048H,2019	1	8	6	82	4
SLL2131F,2019	1	8	2	12	0
SLL3002H,2019	1	8	12	230	2
SLL3131F,2019	1	8	2	9	0
HUB1023S,2019	2	13	7	37	8

Table 56: Course participants for clusters in K-means ($k=2$) for tools visible to students

Site Title	Cluster	Assessment	Communication	Content	Management	Presence
AHS1033F,2019	1	3: 577 : 12.02 : 93.75	2: 940 : 19.58 : 97.92	2: 2880 : 60 : 97.92	1: 98 : 2.04 : 52.08	0: 5527 : 115.15 : 100
AST3002F,2019	1	3: 293 : 17.24 : 88.24	2: 802 : 47.18 : 100	2: 2551 : 150.06 : 100	1: 155 : 9.12 : 94.12	0: 6845 : 402.65 : 100
CML1001L,2019	1	3: 687 : 7.31 : 72.34	2: 1393 : 14.82 : 81.91	2: 2545 : 27.07 : 74.47	1: 421 : 4.48 : 65.96	0: 7460 : 79.36 : 100
CML2001L,2019	1	3: 593 : 8.72 : 76.47	2: 1201 : 17.66 : 91.18	2: 2267 : 33.34 : 89.71	1: 226 : 3.32 : 58.82	0: 4045 : 59.49 : 100
CML4501F,2019	1	3: 1385 : 49.46 : 89.29	2: 146 : 5.21 : 85.71	2: 241 : 8.61 : 92.86	1: 63 : 2.25 : 67.86	0: 1496 : 53.43 : 100
GEO1006S,2019	1	3: 154 : 8.11 : 68.42	2: 75 : 3.95 : 78.95	2: 1506 : 79.26 : 73.68	1: 69 : 3.63 : 68.42	0: 2721 : 143.21 : 100
Greek IB 2019	1	3: 71 : 11.83 : 83.33	2: 6 : 1 : 33.33	2: 0 : 0 : 0	1: 4 : 0.67 : 50	0: 102 : 17 : 100
MEC4105F,2019	1	3: 669 : 74.33 : 100	2: 90 : 10 : 100	2: 938 : 104.22 : 100	1: 72 : 8 : 88.89	0: 1573 : 174.78 : 100
RDL1007F,2019	1	3: 22 : 1 : 50	2: 12 : 0.55 : 31.82	2: 0 : 0 : 0	1: 4 : 0.18 : 9.09	0: 204 : 9.27 : 90.91
SLL1042F,2019	1	3: 285 : 14.25 : 75	2: 147 : 7.35 : 90	2: 69 : 3.45 : 55	1: 93 : 4.65 : 55	0: 549 : 27.45 : 95
SLL1048H,2019	1	3: 1723 : 18.73 : 88.04	2: 540 : 5.87 : 83.7	2: 2444 : 26.57 : 96.74	1: 37 : 0.4 : 18.48	0: 4619 : 50.21 : 94.57
SLL2131F,2019	1	3: 150 : 11.54 : 76.92	2: 36 : 2.77 : 84.62	2: 0 : 0 : 0	1: 30 : 2.31 : 61.54	0: 219 : 16.85 : 100
SLL3002H,2019	1	3: 89 : 0.39 : 23.48	2: 45 : 0.2 : 11.3	2: 0 : 0 : 0	1: 13 : 0.06 : 3.04	0: 405 : 1.76 : 68.7
SLL3131F,2019	1	3: 192 : 21.33 : 100	2: 28 : 3.11 : 100	2: 14 : 1.56 : 33.33	1: 9 : 1 : 33.33	0: 213 : 23.67 : 100
HUB1023S,2019	2	3: 1465 : 32.56 : 93.33	5: 554 : 12.31 : 84.44	4: 4667 : 103.71 : 88.89	1: 48 : 1.07 : 33.33	0: 7743 : 172.07 : 100

Table 57: K-means ($k=2$) event to tool category comparison for tools visible to students

The student interactions for the tool categories in Cluster 1 reveal a wide range of interaction patterns. The values for each tool category show the number of tools, the total number of events for the category, the total average event per student, and the percentage of use of that category by the

students enrolled in the site. Using the assessment tool value for the site “AHS1033F,2019” (the first site and first category entry in Table 57) the value can be interpreted as follows:

3 : 577 : 12.02 : 93.75

- 3 Assessment tools (Assignment, Gradebook, Evaluation).
- 577 total events by students.
- 12.02 average events per student that used tools in this category.
- 93.75% of students enrolled in this site used tools in this category (45 Students enrolled in this site – 42 of which accessed and used tools in this category).

The “Presence” category indicates that a student accessed the site at a point in time, that is why the number of tools for that category is 0, as it does not relate to a specific tool but just the site itself. Ideally, the usage percentage for all sites would be 100%, but as the data shows, this is not always the case. The sites that show less than 100% usage are:

- “RDL1007F, 2019” (90.91%) : 1st-year Private Law course site taught in the 1st Semester with low interactions (total) across all tool categories and sparse content with 22 students enrolled.
- “SLL1042F,2019” (95%): 1st year Afrikaans Intensive course site taught in the 1st Semester with 13 students enrolled, a reasonable amount of content and interaction in all tool categories.
- “SLL1048H,2019” (94.57%): 1st year special course site for Afrikaans for Health and Rehabilitation Sciences. It is a half-course taught over a full year with 82 students enrolled. The site contains content in all the tool categories and a high number of interactions.
- “SLL3002H,2019” (68.7%): a 3rd-year course called “Becoming a Doctor (languages) Part IB”. There are 230 students enrolled, and the structure and content of the site seem to be centred around scheduling and communicating tutorials to the students facilitated by the School of Languages & Literature for the Faculty of Health Sciences. This might show why the interactions are so low, as students can share the announcements, and it’s not necessary to access the site.

The “Content” category for the sites (“Greek IB 2019”, “RDL1007F,2019”, “SLL2131F,2019”, and “SLL3002H,2019”) that don’t have any student interaction (2:0:0:0) contains the overview and resources tool. The overview tool is the landing page for all sites and may contain information about recent announcements, messages, discussion posts, and calendar events. The resources tool allows lecturers to share files with students and is part of the default tools selected on the creation of a course site; in these individual sites, there are no files or folders available to students, which explains why there are no events stored.

Cluster 2 does contain more tools (13) than Cluster 1 (8), but in terms of overall interactions, there does not seem to be a clear pattern. In both clusters, all the sites show a high usage percentage in these categories, indicating that most students are utilising the available tools to interact with these categories. Based on research by Le et al. (2022), Mwalumbwe & Mtebe (2017), and Zhang et al. (2020), higher interactions in important categories such as Assessment, Communication, and Content suggest a positive academic outcome.

The AGNES ($k=2$) (Table 55) algorithm clustered the data set around 2 clusters, 443 and 554 in size (Figure 79). The representative site for the two clusters differ significantly in the number of students enrolled and the number of tools available on each site.

Title	Total Tools	Lecturer	Student	Teaching Support	Other
PBL4505S,2019	12	3	19	1	0
Tests STA1000S,2019	4	4	1226	6	0

Table 58: Course participants for clusters in AGNES ($k=2$) for tools visible to students

Site Title	Cluster	Assessment	Communication	Content	Management	Presence
Tests STA1000S,2019	1	1: 45580 : 36.55 : 96.55	1: 3441 : 2.76 : 76.02	1: 7112 : 5.7 : 94.55	1: 11946 : 9.58 : 96.87	0: 44324 : 35.54 : 100
PBL4505S,2019	2	2: 118 : 6.21 : 100	5: 121 : 6.37 : 94.74	3: 1197 : 63 : 100	2: 96 : 5.05 : 100	0: 2367 : 124.58 : 100

Table 59: AGNES ($k=2$) event to tool category comparison for tools visible to students

For “PBL4505,2019”, the students used all the tools except for communication, which is lower at 94.74%. The usage of the sites also varies greatly. “Tests STA1000S,2019” is used to administer assessments using the “Tests & Quizzes” tool, with the communication tool “Announcements” used to inform students of exam times, signup events, and tutorials. The management tool is added to allow students to sign up for course, practical, and project times. The “Overview” tool, which is categorised as a content tool, shows the welcome message to the site. The site has 129,232 events in its 20-week duration (tracked from first student interaction to last student interaction). It has a lower average event per student (105.41) than “PBL4505,2019”, which has a rate of 211.95 but a much lower total event count of 4027 over its 26 active weeks. For all tool categories, the more populous site “STA” has a lower average event count except, of course, for its main purpose, the assessments (36.55 to 6.21). The representative sites had a high percentage of use across categories, but there was no indicative pattern observed between clusters.

4.2.3.3 Categories of Tools Visible to Students – Clustering

Using the individual tools for analysis would work for this instance of Sakai. However, as each can be configured with different contribution tools (Sakai Contribution Tools) or various LTI added, it’s more appropriate to use the categories of tools. This makes the research comparable to a wider range of LMSs.

	Score	Method	Clusters
Connectivity	12.429762	hierarchical	2
Dunn	0.070252	hierarchical	10
Silhouette	0.543245	hierarchical	2

Table 60: Categories of tools visible to students’ cluster comparison (Internal) – optimal

The clValid recommendation for the categorised tools visible to the student data set recommends the hierarchical clustering (AGNES) in all of the internal validation methods. The number of the clusters differs a little, but viewing the Dunn index graph (Figure 83) gives a slightly clearer picture of the recommendation as the hierarchical clustering (AGNES) starts with a high Dunn index at 2 clusters, drops down at 3 and then increases to 10, which at that point is a larger value than for 2 clusters. Plotting this further also shows that increasing the clusters increases the Dunn index, but as the highest value with the lowest number of clusters is preferable, 2 clusters seem to be the best.

Categorised Tools Visible to Student - Internal Validation

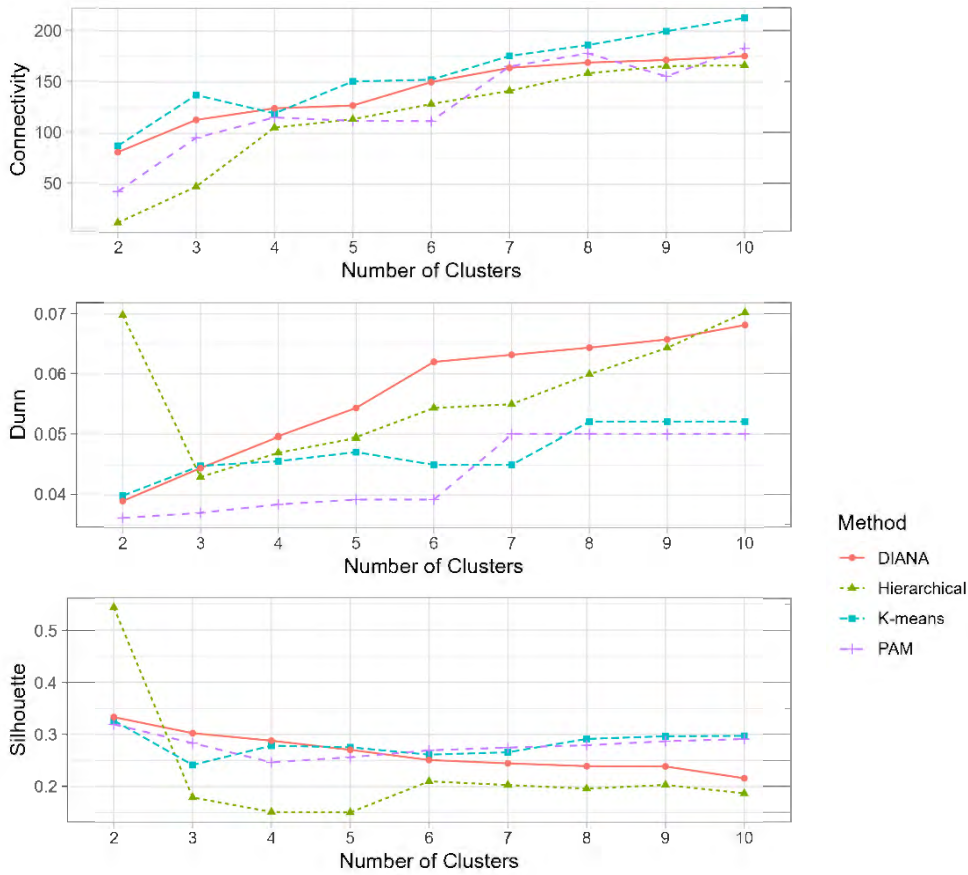


Figure 83: Categories of tools visible to students' cluster comparison (Internal) – Graphs

Plotting the gap statistics, total within sum of squares (WSS), and silhouette calculation for hierarchical clustering (AGNES) on the category of tools visible to students' data shows a small bend in the WSS graph around the 2-cluster point. The silhouette graph also shows 2 as the optimal number of clusters. Both reinforce the results from the cValid internal measurement.

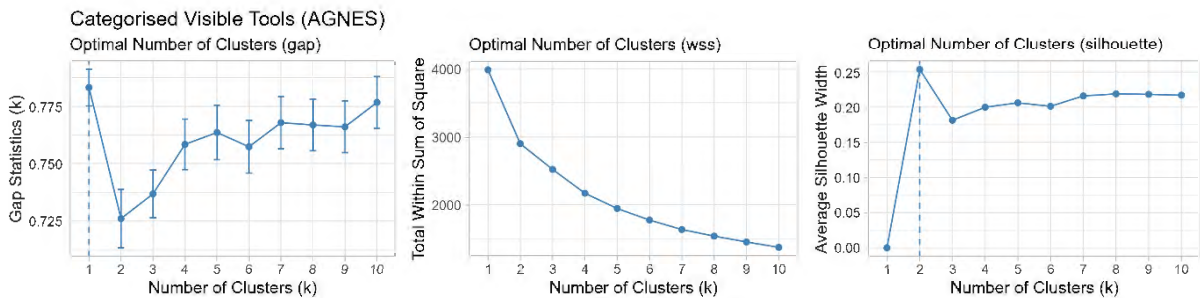


Figure 84: AGNES – Determine the optimal number of clusters using (gap, wss, silhouette) for categorised tools visible to students.

This is borne out by the results from the cluster size graph (Figure 85) and silhouette plot (Figure 86), which give the best result seen so far. The average silhouette width crosses both clusters; the average silhouette score of 0.254 indicates that the clusters are moderately well-separated, and most data points have a moderate degree of conformity to their assigned clusters.

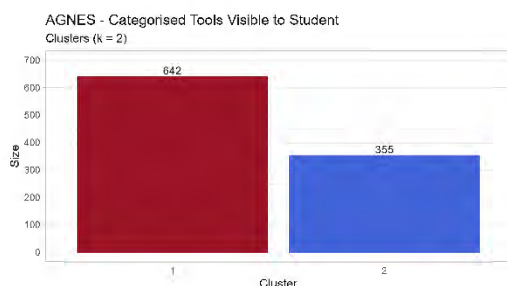


Figure 85: AGNES (k=2) cluster size for categorised tools visible to students

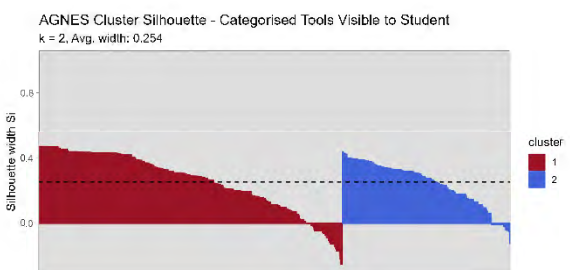


Figure 86: AGNES (k=3) cluster silhouette for categorised tools visible to students

The dendrogram for the hierarchical clustering (AGNES) shows a very nice distinction between the two clusters.

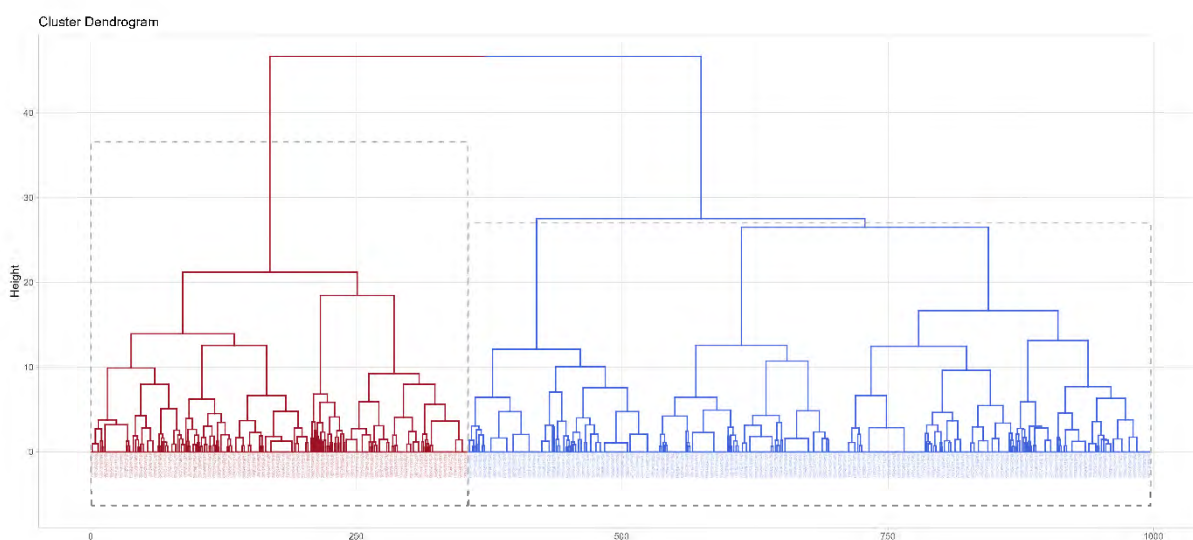


Figure 87: AGNES (k=2) dendrogram for categorised tools visible to students

Cluster	Assessment	Communication	Content	Management
1	0 : 2.646 : 6 (1.015)	0 : 2.813 : 9 (1.371)	0 : 2.852 : 6 (0.816)	0 : 1.316 : 4 (0.587)
2	1 : 3.862 : 7 (0.864)	1 : 4.738 : 11 (1.736)	2 : 3.611 : 7 (0.978)	1 : 2.248 : 4 (0.452)

Table 61: AGNES (k=2) event to tool for categories of tools visible to students

Site Title	Cluster	Assessment	Communication	Content	Management
Accompanying 2,2019	1	3	1	3	1
DOH1002F,2019	2	3	5	4	2

Table 62: AGNES (k=2) sites in the centre of cluster for categorised tools visible to students

Cluster 1 has almost the same mean number of tools in each category as Cluster 2, but the minimum is 0, meaning that some sites in that cluster do not have a tool in one of the categories. Looking at the maximum number of tools on the site, Cluster 2 has a higher maximum in all the categories but more so in Communication. Overall, Cluster 2 contains more tools in each category than Cluster 1.

4.2.3.4 Categories of Tools Visible to Students - Comparing Clusters to Event Data

The observations in the middle of the clusters are “Accompanying 2,2019” for Cluster 1 (low number of tools, and possibly missing in some categories) and “DOH1002F,2019” for Cluster 2 (more tools in each category). Table 63 details the participant data for the two sites, the site with fewer students corresponds to the site with fewer tools available to those students.

Title	Total Tools	Lecturer	Student	Teaching Support	Other
Accompanying 2,2019	8	3	4	0	0
DOH1002F,2019	16	9	206	1	2

Table 63: Course participants for clusters in AGNES ($k=2$) for categorised tools visible to students

The tools in the site for “Accompanying 2,2019” are the default selected ones and, on review of the content, seem to be mainly focused on announcements (13 announcements) for the students. The assignment tool exists but is unused, as are almost all the other tools except resources, which contain a single file (a music piece). The course ran for 43 weeks and only had 121 interactions from students, although the four students in the site did “access” all the teaching and learning tools (100% in Assessment, Communication, and Content)

Site Title	Cluster	Assessment	Communication	Content	Management	Presence
Accompanying 2,2019	1	3: 23 : 5.75 : 100	1: 17 : 4.25 : 100	3: 11 : 2.75 : 100	1: 0 : 0 : 0	0: 65 : 16.25 : 100
DOH1002F,2019	2	3: 8501 : 40.67 : 95.22	5: 11646 : 55.72 : 100	4: 14792 : 70.78 : 99.52	2: 980 : 4.69 : 84.21	0: 20354 : 97.39 : 100

Table 64: AGNES ($k=2$) event to tool category comparison for categorised tools visible to students

The course site for “DOH1002F,2019” has 206 students who access almost all the tools in the course (Assessment: 95.22%, Communication: 100%, Content: 99.52%, Management: 84.21%, Presence: 100%). In the 48 weeks that students accessed the site, they generated 60,330 events, 292.86 events per student, or 6.1 meaningful events per student per week. The tools available to the student include Lecture Recording, a few links to outside websites, the blog tool, and a large collection of files in the resources tool.

Based on usage data, it appears that Cluster 1 is primarily used as a communication channel rather than for blended learning. In contrast, Cluster 2 is a more comprehensive blended learning course site that takes full advantage of available teaching and learning tools.

4.2.4 Summary

In summary, the participants of course sites, the tools visible to students, and the categorised tools visible to students' data sets were clustered using various unsupervised learning algorithms discussed in the methodology chapter. The data sets were also evaluated using the internal validation measurements (“Connectivity”, “Dunn”, and “Silhouette”) defined in the `clValid` package in R (Brock et al., 2008). This helped to identify the optimal number of clusters and the most appropriate clustering algorithm, which was then verified using the gap statistics, total within sum of squares (WSS) and silhouette methods for the chosen clustering algorithm.

The result and importance of the clustering results will be discussed in the next chapter.

5 Discussion

5.1 Introduction

In the previous section, the course site data was analysed in terms of participants, the tools visible to students and the categories for those tools. The initial exploration of each dataset involved the statistical (Hopkins statistics) and visual analysis (VAT) of the clustering tendencies of each data set. This was followed by the four popular unsupervised learning clustering algorithms to determine the representative course sites for the clusters. The aim was to see if certain tool usage patterns could be found which indicated the pedagogical use of the course sites, with the intention to see if sites focused on assessments, communication and collaboration, or content delivery (page, outside content, or downloadable resources).

5.2 Enrolled Students and Events

The participant data set contains the number of users in particular roles enrolled in the course sites (997). The roles are specific to UCT's implementation of Sakai; although there might be similarly named roles, their respective permissions and access might differ. As such, the comparison was done on a combined participant data set, which consist of "Lecturer", "Student", "Teaching Support", and "Other" roles. The Hopkins statistic (0.96889) for this combined data set indicated a high degree of clusterability. The correlation graph showed a slight positive correlation between all the roles and an especially high correlation between "Student" and "Teaching Support" (0.61). This indicates that having more students in a site would necessitate more teaching support; from the full role correlation matrix, this means more tutors in UCT's context. The clusterability of the combined participant data set is confirmed visually with the VAT diagram.

The gap statistics, the total within sum of squares, and the silhouette method were run on the participant data set for each of the four popular clustering algorithms. The cValid internal validation method was used to compare the clustering algorithms and confirm the optimal cluster size choice obtained from the other assessment methods. This resulted in the K-means and AGNES clustering algorithms, with cluster size $k=2$ giving the best results for the participant data set. In both cases, the clusters appear to be subdivided mostly on the number of enrolled users in the course sites, predominantly based on the number of students.

The number of students in a course site directly impacts the number of events generated. The Spearman correlation coefficient of 0.844 indicates a strong positive monotonic relationship between enrolled students and the total number of events. This means that as the number of students increases, the number of events also increases, although not necessarily at a constant rate. The Pearson correlation coefficient of 0.602 indicates a moderate positive linear relationship between the two variables (Events to Students). Pearson assumes a linear relationship, meaning that as the number of students increases, the number of events generated also increases proportionally. Therefore, the increase in the number of students enrolled in a course site increases the chances of generating more events on that site.

5.3 Tools Used in Course Sites

There are 50 tools available to lecturers to include in their course sites, of which 42 are generally made visible to students (internal and LTI), and 32 are active in the events table (internal). The tools used in all the sites were compared to those visible to students, showing that most sites chose the default set of tools. This could indicate that the set of default tools selected matches the

requirements for teaching and learning. Almost half of the sites (456) used lecture recording, thus adding video content to their sites. The tool choices also indicate a long tail of tools (21) used by fewer than 20 sites (2.006%), which begs the question of the support cost for these tools. It became clear that the top tools used in large sites are also the most commonly chosen tools for all sites, after mapping the number of students to the selected tools. It was also observed that sites with fewer students tend to use more non-standard tools. This may suggest that with fewer students, lecturers have the opportunity to interact more and explore different tools to enhance the learning experience.

The correlation between the number of tools added to a course site and the number of students enrolled is relatively weak, as indicated by the Pearson (0.20665) and Spearman (0.20923) coefficients. This suggests that the number of students in a course does not significantly impact the selection of tools in the site. It is more likely that the default set of tools added during site creation, the pedagogical approach of the instructor, the available teaching support, system constraints, and/or time constraints are the factors that determine the selection of tools in a course site.

The Hopkins statistic indicates high clusterability in the data set “tools visible to students” (0.94312) and all tools. The data set chosen to run the cValid internal validation method was for the tools visible to students, as the event data used to compare the tool selection only contains the interactions with students.

Based on cValid’s internal validation method, K-means, AGNES, and DIANA were recommended for clustering with 2 clusters. However, the resulting cluster sizes and silhouette graphs indicated that DIANA did not produce reasonably sized clusters. K-means produced a long list of course sites sharing the closest distance to the cluster's centre. The participants for the first cluster (14-course sites) turned out to be a diverse distribution of a number of students and lecturers with the same tool choices - 3 in Assessment, 2 in Communication, 2 in Content and 1 in Management. The event analysis showed that the students accessed most of the tools in each category (with the exception of “Content” as described in 4.2.3.2 - Tools Visible to Students - Comparing Clusters to Event Data). This diverse usage of tools in the site might indicate that the teaching and learning approach for the subject matter might have more impact on the interactions than just the choice of tools.

The selection of AGNES with 2 clusters showed that the clusters are primarily subdivided into sites with a low number of tools and sites with a high number of tools in the varying categories. The event comparison showed that for each representative site in the cluster, a very high percentage of students used all the tools available to them. Cluster 1 was represented by “PBL4505S,2019”, which has only 19 students enrolled and a total event count of 4,027 (average number of events per user of 211.95). The site contains 12 tools focused on the content and communication categories. Cluster 2 is represented by “Tests STA1000S,2019”, which has 1226 students enrolled - 129,232 total events at 105.41 average events per student. This site only contains 4 tools visible to students focused on communication and assessment. In the case of this site, the site with fewer tools had a higher student count and, thus, more events.

Taking the existing list of tools and combining them into categories makes it easier to do cross LMS comparisons. This data set is called “categorised tools visible to students”, and the clustering method recommended for this data set is hierarchical clustering (AGNES) with a cluster size of 2, which was supported by the cluster size and silhouette graphs. The corresponding representative sites confirmed previous results where students accessed almost all the tools in the sites. The example sites in this cluster followed the correlation that more tools resulted in more interactions (Figure 82).

5.4 Student Interactions

The representative sites chosen for the clusters in the “Tools Visible to Students” and the categorised version of that data showed that a high percentage of enrolled students accessed all the tools in a category. The variation in the number of events per category and total percentage of use across all the representative sites indicate that the tool selection might not be as important as the teaching approach in determining student interaction (surmised as there is no data on the teaching approach).

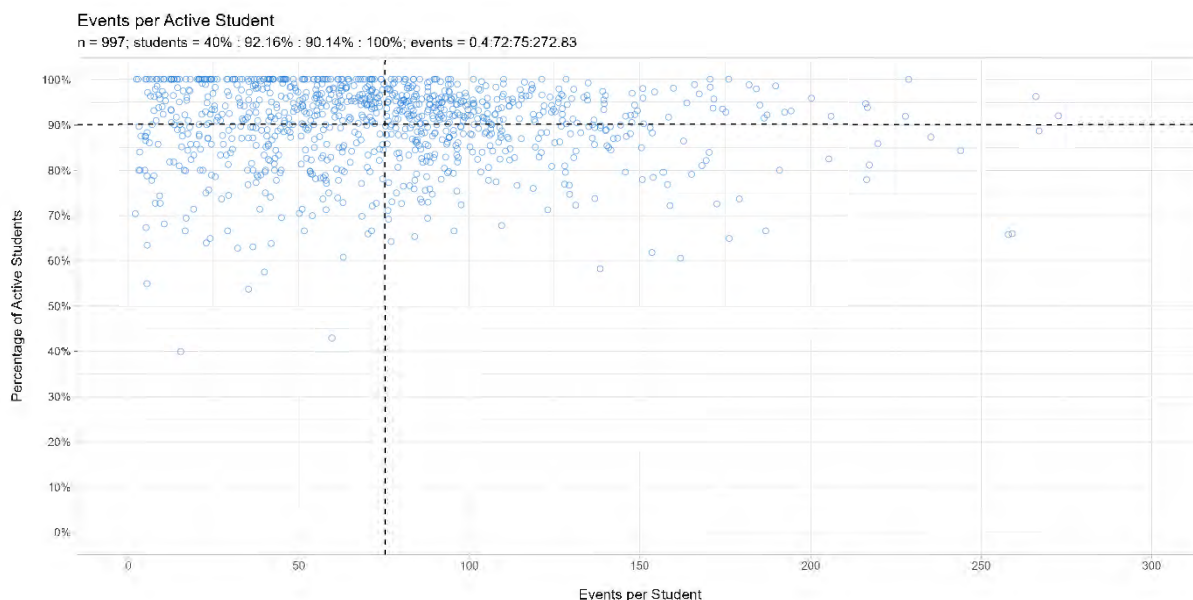


Figure 88: Average percentage of active students and events per student (min:median:mean:max)

Figure 88 illustrates the percentage of students who accessed tools on a course site and the average number of events for those students. This shows that around 90% of students access a course site, creating around 75 interactions on average. This reinforces the findings from the representative sites, which (Table 59, Table 61, and Table 66) generally indicate that most students access the site and interact with most of the tools on the site.

This corresponds nicely with the findings in the literature review, which indicate that interactions with the LMS have a positive impact on student performance (Le et al., 2022; Mwalumbwe & Mtebe, 2017). This ties into the findings from (Zhang et al., 2020), where they found that interaction with the course site, particularly opening files (accessing content tools), positively correlates with a higher grade. This was also confirmed by (Fisher & Parker, 2016), where in their analysis, half the courses showed a significant positive relationship between online learning resource utilisation frequency (i.e., course site visits and online learning resource downloads) to achieved course grade.

6 Conclusion

6.1 Overview

This research paper aims to examine the tools utilized in blended and online courses at the University of Cape Town. It evaluates the students' interactions in the undergraduate course sites for the year 2019 and compares the tools used to the categories of tools identified in the literature review.

To this aim, the six chapters in this study cover the Learning Management System (LMS) and its current role in higher education and the general marketplace. The tools available to users of an LMS are explored and categorised. Four popular unsupervised learning algorithms (K-means, PAM, AGNES, DIANA) are introduced and discussed in order to cluster the structure of the course sites and compare the tool selection to student interactions. The data sets obtained were clustered using the methods discussed, and the resulting tool selection and event data of the representative sites were evaluated.

6.2 Research Objectives

Literature Review

1. Examine the Learning Management System (LMS) marketplace.

The Learning Management System (LMS) marketplace was examined, and the current market leaders in the software architectural groups (commercial or non-commercial, with closed, open, or partially open) were identified. The features and tools available to the various systems were explored; this led to identifying categories of tools available across all LMSs and used in the analysis phase of the tools visible to students.

2. Investigate the importance of LMSs in higher education.

The literature review also indicated that an LMS is of the highest importance for higher education institutions and has positively impacted the traditional teaching-learning process. LMSs are widely used in higher education institutions globally and are considered one of the most important instructional technologies for academic success. They also have a transformative effect on traditional teaching and learning, facilitate flexible classroom management, and provide the faculty with a flexible way of teaching and grading students' assessments.

3. Research the impact of student interactions with tools on learning outcome results.

Furthermore, the interactions with the LMS indicated that for certain cases pertaining to a specific category of tools, the more a student engages with the tools in the LMS, the higher the academic outcome is. There are some limitations of LMSs that centre around the architectural design of the software, the interaction with students and the tool support for higher-order thinking, or lack thereof. The scope of previous studies was limited to either a set of closely related courses or introducing a new tool and its impact on student behaviour and performance. This study expanded that scope to encompass more course sites.

Data and Analysis

4. Collect data from Sakai (LMS) pertaining to course structure and student interactions.

The data set selected for this study pertains to 2019 undergraduate course sites. The selection of the data was influenced by the #FeesMustFall and COVID-19 nationwide lockdown, which changed the student's interaction with the LMS. The site structure for the course sites was initially extracted directly from the Sakai database, and later, the REST API was used to obtain the tool configurations and access based on the student role permissions. The site data was analysed to extract the blended and online undergraduate courses and remove duplicates. The event data for the constrained data set was extracted from the Sakai event and session tables.

5. Analyse the course structure and classify.

The initial set of enrolled users in the course sites was used to evaluate and compare the chosen unsupervised learning algorithms. This helped in understanding the relationship between students, tools, and events. The analysis revealed that having more enrolled students might result in more events, but it is not necessarily a linear correlation. The correlation still depends on the tool choices. The algorithms successfully created reasonable clusters, in this case 2, based on the number of enrolled users, primarily the number of students.

Although the number of available tools is similar across all categories, the category of course development and delivery generated more events than all other categories combined. This is despite having one of the lowest numbers of event types for students. These findings suggest that the LMS is used as a knowledge repository in which lecturers scaffold content delivery for most sites. Some representative sites in the analysis are exceptions, as they are only used for communication or assessment.

The analysis of the tools visible to students generally also produced 2 clusters, defined by the total number of tools. This indicates that the clustering used in this study is likely not the best approach to analyse the structure of course sites, or the data set, although highly clusterable, needed more exploration. The comparison of the representative tools in the clusters highlighted the correlation that more students in the site generally led to a higher number of events, and in some cases, that could be translated to a higher number of tools leading to more student interactions for particular categories of tools. The result for K-means clustering, 2 clusters, showed that the event interaction is more complex than the tool choices for the site and depends on the pedagogical approach of the particular course and the interaction of the lecturer and teaching support. This highlights the gap between the tools available and what the lecturer might want to achieve on the site.

After analysing the categories of tools available to students, a hierarchical clustering algorithm (AGNES) was applied, which resulted in two clusters. The clusters were defined based on the total number of tools in each category. Cluster 1 had a low total tool count, while Cluster 2 had a higher total tool count. As with the previous findings, more tools and students resulted in more events per category.

6. Compare student activity based on the classification of the course structure.

Finally, when reviewing the usage percentage of the tools in each category, it was found that most students did utilise the LMS and access a high percentage of tools in each category. In the scope of the events generated by the tools and analysed for these course sites, the conclusion is that a good number of students would successfully achieve a passing grade for their enrolled courses.

It should be noted that the analysis conducted had certain limitations regarding the events tracked by the system. Additionally, it assumes a one-sided perspective, as it only considers the interactions between the students and the online learning system.

6.3 Future Work

Clustering course sites on participants, "the tools visible to students", and "categorised tools visible to students" could have yielded better clusters using different clustering algorithms. Most clusters were distinguished by the number of tools included in the site, and most sites used the default selected tool set. This could be because the default selected tools, customised in Vula for UCT, met all the lecturers' requirements.

The unsupervised clustering algorithms might not be best suited to cluster the site structure, and further investigation into clustering sites into functional groups might inform a better approach.

The influence of pedagogical teaching strategies would also have an influence on tool choices, so a further avenue of investigation would be the discussion with lecturers (for specific courses) to determine their approaches and the tools that support their teaching model.

It might also be interesting to approach the site structure from the student interaction perspective. It is possible to determine the preferred tool interactions by observing how students approach the learning process. This can help identify which tools support teaching and learning. This could also be applied to the interactions for other roles like "Lecturer", "Support Staff", and "Tutor".

It would also be interesting to investigate the tool choices between undergraduate and post-graduate course sites and the difference in learning approaches that this provides to lecturers.

Finally, the information on participants and the relationship between tool selection (site structure) and student interactions did indicate that there is an influential relationship between them, and the additional information may inform further studies on practical tool usage and improvements in LMSs.

References

Literature

- Ahmed, K., & Mesonovich, M. (2019). Learning management systems and student performance. *International Journal of Sustainable Energy Development (IJSED)*, 7(1), 582–591. <https://www.academia.edu/download/66835157/ijels.2046.4568.2019.pdf>
- Al-Rahmi, W. M., Othman, M. S., Yusof, L. M., & Musa, M. A. (2015). Using social media as a tool for improving academic performance through collaborative learning in Malaysian higher education. *Review of European Studies*, 7(3), 265–275. https://heinonline.org/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/rveurost7§ion=78
- Amathuba. (2023). <https://cilt.uct.ac.za/support-and-services/amathuba>
- Amponsah, S., Badu-Nyarko, S. K., Obodai, G. A. N. S., & Anane, P. (2019). Learning environment for supporting undergraduate online distance education students. *Student Support toward Self-Directed Learning in Open and Distributed Environments*, 78–102. <https://doi.org/10.4018/978-1-5225-9316-4.ch004>
- Apereo. (2023a). *Sakai Contrib Tools*. <https://github.com/sakaicontrib/>
- Apereo. (2023b). *Sakai LMS Community*. <https://www.sakailms.org/sakai-lms-community>
- Asiri, M. J. S., Mahmud, R. B., Bakar, K. A., Fauzi, A., & Ayub, M. (2012). Factors Influencing the Use of Learning Management System in Saudi Arabian Higher Education: A Theoretical Framework. *Higher Education Studies*, 2(2), 125. <https://doi.org/10.5539/HES.V2N2P125>
- Avci, Ü., & Ergün, E. (2019). Online students' LMS activities and their effect on engagement, information literacy and academic performance. *Interactive Learning Environments*, 30(1), 71–84. <https://doi.org/10.1080/10494820.2019.1636088>
- Best Education Learning Management (ELMS) Systems 2023*. (2023). SoftwareReviews Website. <https://www.softwarereviews.com/categories/education-learning-management>
- Bezdek, J. C., & Hathaway, R. J. (2002). VAT: A tool for visual assessment of (cluster) tendency. *Proceedings of the International Joint Conference on Neural Networks*, 3, 2225–2230. <https://doi.org/10.1109/IJCNN.2002.1007487>
- Blackboard (Anthology)*. (2023). <https://www.blackboard.com/en-mea/teaching-learning/learning-management>
- Blackboard Tools*. (2023). <https://help.blackboard.com/Learn/Instructor/Ultra>
- Brightspace (D2L)*. (2023). <https://www.d2l.com/brightspace/>
- Brightspace Tools*. (2023, September 9). <https://community.d2l.com/brightspace/kb/categories/1334-overview-of-your-course-tools>
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). clValid: An R Package for Cluster Validation. *Journal of Statistical Software*, 25(4), 1–22. <https://doi.org/10.18637/JSS.V025.I04>
- Brooks, D. C., & Pomerantz, J. (2016). *ECAR study of undergraduate students and information technology, 2016*. <https://library.educause.edu/resources/2016/6/2016-students-and-technology-research-study>

- Brown, M., Millichap, N., & Dehoney, J. (2015, June 22). *What's Next for the LMS?* Educause Review. <https://er.educause.edu/articles/2015/6/whats-next-for-the-lms>
- Burgos, D. (2019). Background Similarities as a Way to Predict Students' Behaviour. *Sustainability*, 11(24), 6883. <https://doi.org/10.3390/su11246883>
- Camirero, A., & Hernández, R. (2013). Choosing the right LMS: A performance evaluation of three open-source LMS. *2013 IEEE Global Engineering Education Conference (EDUCON)*. <https://ieeexplore.ieee.org/abstract/document/6530119/>
- Canvas (Instructure). (2023). <https://www.instructure.com/canvas>
- Canvas Tools. (2023). <https://learn.canvas.cornell.edu/canvas-tools/>
- Cavus, N., & Alhih, M. S. (2014). Learning management systems use in science education. *Procedia-Social and Behavioral Sciences*, 517–520. <https://www.sciencedirect.com/science/article/pii/S1877042814043572>
- Cavus, N., & Zabadi, T. (2014). A comparison of open source learning management systems. *Procedia-Social and Behavioral Sciences*, 143, 521–526. <https://www.sciencedirect.com/science/article/pii/S1877042814043584>
- Cini, L. (2019). Disrupting the neoliberal university in South Africa: The# FeesMustFall movement in 2015. *Current Sociology*, 67(7), 942–959. <https://doi.org/10.1177/0011392119865766>
- Cucu, C. (2014). BLENDED LEARNING USING MOODLE AT THE „1 DECEMBRIE UNIVERSITY" IN ROMANIA. *Annals of the University of Petroşani. Economics*, 14(1), 41–48. https://www.upet.ro/annals/economics/pdf/2014/part1/Cucu_C.pdf
- Czerniewicz, L., Ravjee, N., & Mlitwa, N. (2007). Mapping the landscape ICTs in Higher Education in South Africa. *Report for the Council for Higher Education*.
- Davidson, B. (2022). *Data Analytics Living Textbook*. https://bookdown.org/brittany_davidson1993/bookdown-demo/
- Dhiman, S. B., & Gera, M. (2021). Enhancing academic achievement of commerce students through blended e-learning. *International Journal of Research -GRANTHAALAYAH*, 9(7), 137–144. <https://doi.org/10.29121/granthaalayah.v9.i7.2021.4097>
- EDUCAUSE Learning Initiative. (2015). *7 THINGS YOU SHOULD KNOW ABOUT NGDLE*. <https://library.educause.edu/~media/files/library/2015/12/eli7127-pdf.pdf>
- Falvo, D. A., & Johnson, B. F. (2007). The use of learning management systems in the United States. *TechTrends*, 51(2), 40–45. <https://search.proquest.com/openview/070d25240882e6facbc54d498b4cb395/1?pq-origsite=gscholar&cbl=40581>
- Firdiyewek, Y. (1999). Web-Based Courseware Tools: Where Is the Pedagogy? *Educational Technology*, 39(1), 29–34. <https://www-jstor-org.ezproxy.uct.ac.za/stable/44429009>
- Fisher, P. J., & Parker, H. (2016). *An Evaluation of Student Academic Achievement Based On Online Academic Resource Utilisation Frequency and Timeliness: An Exploratory Meta-Analysis in Learning Analytics*. University of Cape Town.

- Freed, N., Klensin, J. C., & Hansen, T. (2015, October 14). *RFC 6838 - Media Type Specifications*. IETF. <https://datatracker.ietf.org/doc/rfc6838/>
- Garrison, D. R., & Kanuka, H. (2004). Blended learning: Uncovering its transformative potential in higher education. *The Internet and Higher Education*, 7(2), 95–105. <https://www.sciencedirect.com/science/article/pii/S1096751604000156>
- Graham, C. R. (2006). Blended learning systems. *The Handbook of Blended Learning: Global Perspectives, Local Designs*, 1, 3–21. <https://books.google.com/books?hl=en&lr=&id=tKdyCwAAQBAJ&oi=fnd&pg=RA1-PA3&dq=what+is+blended+learning&ots=BipKJrzFdl&sig=E2uRAZTXLJf86AJY5miXUS8X0Rw>
- Guo, Y. (2013). ReCa: A Social Relationship Mining Tool in Virtual Learning Environment. *International Journal of Emerging Technologies in Learning (Online)*, 8(3), 13. <https://doi.org/10.3991/ijet.v8i3.2686>
- Helm, M. (2021). *A deep dive into k-means*. <https://towardsdatascience.com/a-deep-dive-into-k-means-f9a1ef2490f8>
- Hill, P. (2022, February 1). *State of Higher Ed LMS Market for US and Canada: Year-End 2021 Edition*. On EdTech Newsletter. https://onedtech.philhillaa.com/p/state-of-higher-ed-lms-market-for-us-and-canada-year-end-2021-edition/?utm_source=onedtech.philhillaa.com&utm_medium=referral&utm_campaign=state-of-higher-ed-lms-market-for-us-and-canada-year-end-2021-edition
- Hodges, W., Yannick, L., & Jones, M. (2023). *Sakai Release Date list*. Confluence. <https://sakaiproject.atlassian.net/wiki/spaces/DOC/pages/17281679656/Sakai+Release+Date+list>
- Holmes, K. A., & Prieto-Rodriguez, E. (2018). Student and staff perceptions of a learning management system for blended learning in teacher education. *Australian Journal of Teacher Education (Online)*, 43(3), 21–34. <https://search.informit.org/doi/abs/10.3316/informit.477648976239419>
- Hrastinski, S. (2019). What Do We Mean by Blended Learning? *TechTrends*, 63(5), 564–569. <https://doi.org/10.1007/S11528-019-00375-5>
- Kassambara, A. (2023a). *Agglomerative Hierarchical Clustering*. Datanovia. <https://www.datanovia.com/en/lessons/agglomerative-hierarchical-clustering/>
- Kassambara, A. (2023b). *Assessing Clustering Tendency*. <https://www.datanovia.com/en/lessons/assessing-clustering-tendency/>
- Kassambara, A. (2023c). *Cluster Validation Statistics: Must Know Methods*. Datanovia. <https://www.datanovia.com/en/lessons/cluster-validation-statistics-must-know-methods/>
- Kassambara, A. (2023d). *Divisive Hierarchical Clustering*. Datanovia. <https://www.datanovia.com/en/lessons/divisive-hierarchical-clustering/>
- Kassambara, A. (2023e). *Hierarchical Clustering in R: The Essentials*. Datanovia. <https://www.datanovia.com/en/courses/hierarchical-clustering-in-r-the-essentials/>

- Kaufman, L., & Rousseeuw, P. J. (1990). Clustering Large Applications (Program CLARA). In *Wiley Series in Probability and Statistics* (pp. 126–163). Wiley.
<https://doi.org/10.1002/9780470316801.ch3>
- Keita, Z. (2023). *Principal Component Analysis (PCA) in R Tutorial*.
<https://www.datacamp.com/tutorial/pca-analysis-r>
- Kerr, S. (2011). High School Online: Pedagogy, Preferences, and Practices of Three Online Teachers. *Journal of Educational Technology Systems, 39*(3), 221–244. <https://doi.org/10.2190/ET.39.3.B>
- Kyndt, E., & Baert, H. (2013). Antecedents of employees' involvement in work-related learning: A systematic review. *Review of Educational Research, 83*(2), 273–313.
<https://doi.org/10.3102/0034654313478021>
- Landa, N., Zhou, S., & Marongwe, N. (2021). Education in emergencies: Lessons from COVID-19 in South Africa. *International Review of Education, 67*(1–2), 167–183.
<https://doi.org/10.1007/S11159-021-09903-Z>
- Langa, M., Ndelu, S., Edwin, Y., & Vilakazi, M. (2017). # Hashtag: An analysis of the # FeesMustFall movement at South African universities. *# Hashtag: An Analysis of The # FeesMustFall Movement at South African Universities*. <https://www.africaportal.org/publications/hashtag-an-analysis-of-the-feesmustfall-movement-at-south-african-universities/>
- Lawson, R. G., & Jurs, P. C. (1990). New index for clustering tendency and its application to chemical problems. *J. Chem. Inf. Comput. Sci., 30*(1), 36–41. <https://doi.org/10.1021/Ci00065A010>
- Le, B., Lawrie, G. A., & Wang, J. T. H. (2022). Student Self-perception on Digital Literacy in STEM Blended Learning Environments. *Journal of Science Education and Technology, 31*(3), 303–321.
<https://doi.org/10.1007/S10956-022-09956-1>
- Learning Management System Market Size & Global Forecast*. (2023). Market and Markets.
<https://www.marketsandmarkets.com/Market-Reports/learning-management-systems-market-1266.html>
- Learning Platforms Update Project*. (2023). <https://cilt.uct.ac.za/projects/learning-platforms-update-project>
- Lee, J. E., Recker, M., Bowers, A. J., & Yuan, M. (2016). Hierarchical Cluster Analysis Heatmaps and Pattern Analysis: An Approach for Visualizing Learning Management System Interaction Data. *EDM, 603–604*. <https://doi.org/10.1198/tas.2009.0033>
- LMS Data – Spring 2021 Updates*. (2021). Edutechnica. <https://edutechnica.com/2021/06/21/lms-data-spring-2021-updates/>
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2022). *cluster: Cluster Analysis Basics and Extensions*. <https://CRAN.R-project.org/package=cluster>
- Mehta, A., & Kalyvaki, M. (2017). Learning Management System: Education Research in the Era of Technology. In *Computer-Aided Data Analysis in Chemical Education Research (CADACER): Advances and Avenues* (Vol. 1260, pp. 9–19). American Chemical Society.
<https://doi.org/doi:10.1021/bk-2017-1260.ch002>

- Mhlanga, D., & Moloi, T. (2020). COVID-19 and the digital transformation of education: What are we learning on 4IR in South Africa? *Education Sciences*, 10(7), 180. <https://doi.org/10.3390/educsci10070180>
- Mlitwa, N., Van Belle, J.-P., & Madhusudhan, M. (2009). The use of ICT for Teaching and Learning in South African Higher Education Institutions. *Conference on Information Science Technology and Management*, 69. https://www.researchgate.net/profile/Nhlanhla-Mlitwa/publication/242170671_The_use_of_ICT_for_Teaching_and_Learning_in_South_African_Higher_Education_Institutions/links/559178b508aed6ec4bf83af1/The-use-of-ICT-for-Teaching-and-Learning-in-South-African-Higher-Education-Institutions.pdf
- Moodle LMS. (2023). <https://moodle.org/>
- Moodle Tools. (2023). https://docs.moodle.org/402/en/Learning_Tools
- Moore, M. G. (1989). Editorial: Three Types of Interaction. *American Journal of Distance Education*, 3(2), 1–7.
- Mwalumbwe, I., & Mtebe, J. S. (2017). Using learning analytics to predict students' performance in Moodle learning management system: A case of Mbeya University of Science and Technology. *The Electronic Journal of Information Systems in Developing Countries*, 79(1), 1–13. <https://doi.org/10.1002/j.1681-4835.2017.tb00577.x>
- O'Brien, L. (2020, April 23). *SoftwareReviews names Sakai, Canvas, and Moodle among Education*. SoftwareReviews - Press Release. <https://www.softwarereviews.com/press-releases/emotional-footprint-awards-2020-education-learning-management-systems>
- Opencast – The free and open source solution for automated video capture and distribution at scale. (2023). <https://opencast.org/>
- Open-core model - Wikipedia. (2023). https://en.wikipedia.org/wiki/Open-core_model
- Parker, S., & Philpot, T. (2023). *Tsugi - Mec Movies*. <https://tsugi-tools.uct.ac.za/store/test/MecMoviesTsugi>
- Pireva, K., Imran, A., & Dalipi, F. (2015). User behaviour analysis on LMS and MOOC. *2015 IEEE Conference on E-Learning, e-Management and e-Services (IC3e)*, 21–26. <https://doi.org/10.1109/IC3e.2015.7403480>
- Reid, L., & Reid, L. (2019). Learning management systems: The game changer for traditional teaching and learning at adult and higher education institutions. *Global Journal of Human Social Science: G Linguistics & Education*, 19(6), 1–13. <https://doi.org/10.34257/GJHSSGVOL19IS6PG1>
- Reyes, N. R., Candeas, P. V., Galan, S. G., Viciano, R., Canadas, F., & Reche, P. (2009). Comparing open-source e-learning platforms from adaptivity point of view. *2009 EAEEIE Annual Conference*, 1–6. <https://doi.org/10.1109/EAEEIE.2009.5335482>
- Rhode, J., Richter, S., Gowen, P., & Miller, T. (2017). Understanding faculty use of the learning management system. *Online Learning*, 21(3), 68–86. <https://doi.org/10.24059/olj.v%vi%i.1217>
- Roderick, M., & Engel, M. (2007). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23(3), 197–227. <https://doi.org/10.1086/518490>

- Sáiz-Manzanares, M. C., Marticorena-Sánchez, R., Díez-Pastor, J. F., & García-Osorio, C. I. (2019). Does the use of learning management systems with hypermedia mean improved student learning outcomes? *Frontiers in Psychology, 10*(FEB), 353851. <https://doi.org/10.3389/FPSYG.2019.00088/BIBTEX>
- Sakai 23 User Guide*. (2023). https://sakai.screenstepslive.com/s/sakai_help/m/114554
- Sakai Features*. (2023). <https://www.sakailms.org/sakai-lms-features>
- Sakai LMS*. (2023). <https://www.sakailms.org/>
- Schubert, E., & Rousseeuw, P. J. (2019). Faster k-Medoids Clustering: Improving the PAM, CLARA, and CLARANS Algorithms. *SISAP 2019, 11807 LNCS*, 171–187. https://doi.org/10.1007/978-3-030-32047-8_16
- Severance, C. (2023). *Tsugi*. <https://www.tsugi.org/>
- Severance, C. R. (2013). *Sakai: Building an Open Source Community*. 235. <http://www.dr-chuck.com/sakai-book/>
- Sinaga, K., & Yang, M. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*. <https://ieeexplore.ieee.org/abstract/document/9072123/>
- The University of Cape Town Selects D2L To Enhance and Future Proof Its Digital Learning Infrastructure*. (2022, June 22). Press Release. <https://www.realwire.com/releases/the-university-of-cape-town-selects-d2l>
- Tsugi - Student Notes*. (2023). <https://tsugi-tools.uct.ac.za/store/test/studentnotes>
- Various. (2019). *UCT Teaching and Learning Report*. https://faculty.uct.ac.za/sites/default/files/content_migration/uct_ac_za/39/files/2019%252520UCT%252520TL%252520Report%252520Approved%252520June%2525202021.pdf
- Vision 2030*. (2023). <https://uct.ac.za/transformation/vision-2030>
- Wan, H., Yu, Q., Ding, J., & Liu, K. (2018). Students' behavior analysis under the Sakai LMS. *Proceedings of 2017 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2017, 2018-January*, 250–255. <https://doi.org/10.1109/TALE.2017.8252342>
- What is the Lessons tool?* (2023). Sakai Community Documentation. https://sakai.screenstepslive.com/s/sakai_help/m/114554/l/1629296-what-is-the-lessons-tool
- What is the Resources tool?* (2023). Sakai Community Documentation. https://sakai.screenstepslive.com/s/sakai_help/m/114554/l/1629391-what-is-the-resources-tool
- Yakubu, M. N., Dasuki, S. I., Abubakar, M., & Kah, M. M. (2020). Determinants of learning management systems adoption in Nigeria: A hybrid SEM and artificial neural network approach. *Education and Information Technologies, 25*(1), 3515–3539. <https://doi.org/10.1007/s10639-020-10110-w>
- Zhang, Y., Ghandour, A., & Shestak, V. (2020). *Using learning analytics to predict students performance in moodle LMS*. <https://digitallibrary.aau.ac.ae/handle/123456789/434>

R

- Auguie, B. (2017). *gridExtra: Miscellaneous Functions for "Grid" Graphics*. <https://CRAN.R-project.org/package=gridExtra>
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2021). *clValid: Validation of Clustering Results*. <https://CRAN.R-project.org/package=clValid>
- Chang, W. (2022). *webshot: Take Screenshots of Web Pages*. <https://CRAN.R-project.org/package=webshot>
- Cheng, J., Sievert, C., Schloerke, B., Chang, W., Xie, Y., & Allen, J. (2023). *htmltools: Tools for HTML*. <https://CRAN.R-project.org/package=htmltools>
- Couture-Beil, A. (2022). *rjson: JSON for R*. <https://CRAN.R-project.org/package=rjson>
- de Vries, A., & Ripley, B. D. (2022). *ggdendro: Create Dendrograms and Tree Diagrams Using "ggplot2."* <https://CRAN.R-project.org/package=ggdendro>
- Galili, T., & Jefferis, G. (2023). *dendextend: Extending "dendrogram" Functionality in R*. <https://CRAN.R-project.org/package=dendextend>
- Hennig, C. (2023). *fpc: Flexible Procedures for Clustering*. <https://CRAN.R-project.org/package=fpc>
- Husson, F., Josse, J., Le, S., & Mazet, J. (2023). *FactoMineR: Multivariate Exploratory Data Analysis and Data Mining*. <https://CRAN.R-project.org/package=FactoMineR>
- Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. <http://www.sthda.com/english/rpkgs/factoextra>
- Maechler, M., Rousseeuw, P., Struyf, A., & Hubert, M. (2022). *cluster: "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.* <https://CRAN.R-project.org/package=cluster>
- Müller, K., & Wickham, H. (2023). *tibble: Simple Data Frames*. <https://CRAN.R-project.org/package=tibble>
- R Core Team. (2023). *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/>
- Ren, K., & Russell, K. (2021). *formattable: Create "Formattable" Data Structures*. <https://CRAN.R-project.org/package=formattable>
- Spinu, V., Grolemond, G., & Wickham, H. (2023). *lubridate: Make Dealing with Dates a Little Easier*. <https://CRAN.R-project.org/package=lubridate>
- Stanley, J., & Arendt, C. (2023). *tidyjson: Tidy Complex "JSON."* <https://CRAN.R-project.org/package=tidyjson>
- Templ, M., Kowarik, A., Alfons, A., de Cillia, G., & Rannetbauer, W. (2022). *VIM: Visualization and Imputation of Missing Values*. <https://CRAN.R-project.org/package=VIM>
- Tiefenbach, T. (2023). *dplyover: Create columns by applying functions to vectors and/or columns in "dplyr."* <https://github.com/TimTeaFan/dplyover>
- Vaidyanathan, R., Xie, Y., Allaire, J. J., Cheng, J., Sievert, C., & Russell, K. (2023). *htmlwidgets: HTML Widgets for R*. <https://CRAN.R-project.org/package=htmlwidgets>

- Wickham, H. (2020). *reshape2: Flexibly Reshape Data: A Reboot of the Reshape Package*. <https://CRAN.R-project.org/package=reshape2>
- Wickham, H. (2022). *stringr: Simple, Consistent Wrappers for Common String Operations*. <https://CRAN.R-project.org/package=stringr>
- Wickham, H. (2023). *tidyverse: Easily Install and Load the "Tidyverse."* <https://CRAN.R-project.org/package=tidyverse>
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., Woo, K., Yutani, H., & Dunnington, D. (2023). *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>
- Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>
- Wickham, H., & Henry, L. (2023). *purrr: Functional Programming Tools*. <https://CRAN.R-project.org/package=purrr>
- Wickham, H., Hester, J., & Ooms, J. (2023). *xml2: Parse XML*. <https://CRAN.R-project.org/package=xml2>
- Wickham, H., Vaughan, D., & Girlich, M. (2023). *tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr>
- Wright, K. (2022). *hopkins: Calculate Hopkins Statistic for Clustering*. <https://CRAN.R-project.org/package=hopkins>
- Xie, Y. (2023). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. <https://CRAN.R-project.org/package=knitr>
- Zhu, H. (2021). *kableExtra: Construct Complex Table with "kable" and Pipe Syntax*. <https://CRAN.R-project.org/package=kableExtra>

Appendix

Appendix – Sakai Tools

The Sakai LMS uses quite a variety of tools; due to how long it's been around, the open-source nature allows for the creation of additional tools and the extensive collaboration and integration capabilities of the system. This appendix describes the categorisation of these tools around the event IDs to increase the reproducibility of the research to other LMS's and to ease the discussion of the classifications of tools in this study.

Categorising the tools used in the LMS has been done in other studies; for instance, in (Caminero & Hernández, 2013) they group the tools in the LMS by video-conferencing, forums or email, evaluation tools such as questionnaires, or grading tools. In (Kerr, 2011) the main categories are content, communication, and management tools. In their study gets elaborated upon during each use-case as shown in the tables below:

Content Tools	Communication Tools
Online Text	Blog
Supplemental websites	Discussion Board
Video clips	Elluminate Live
Audio Clips	

Table 65: Tools Used at Crestview Online (Table 1 - Kerr, 2011)

Communication	Content	Management system
Elluminate Live	Textbook	Project Foundry
Instant Messenger	Websites	
E-mail	Community Expert	

Table 66: Tools Used at Morris Online Academy (Table 2 - Kerr, 2011)

Communication	Content	Management system
Discussion Board	Reading Packet	Events Calendar
E-Mail	Online Text	Checklist
Online Pager	Supplemental Websites	Desire2Learn

Table 67: Tools Used in MRVC History Course (Table 3 - Kerr, 2011)

In the one section (Kerr, 2011) breaks down the tools as courseware designing tools that enable instructors to hide or make certain course features available, synchronous communication (e.g., video, audio, chat rooms, whiteboard), asynchronous communication (threaded discussion board), assessment (e.g., quiz, test, or surveys), course management (calendars, student statistics, grade books), student authoring, and scaffolding tools (advance organisers and annotations).

The next section goes through most of the tools currently available in Sakai and defined on their feature website (*Sakai Features*, 2023), adds the id used in the analysis and description where missing.

Communications and Collaboration

Tool	ID	Description
Announcements	announcement	Post current, time-critical information for your site and schedule notices to display at specific times throughout the term.
Calendar	calendar	Schedule and maintain deadlines, activities and events related to a course or project, link to announcements, assignments, assessments, materials, etc.
Chat	chat	Engage in real-time, text-based conversations with course or project participants.
Commons	commons	Interact with site participants in a Facebook-inspired, asynchronous social messaging tool.
Contact Us	contact	Quickly and easily direct questions and feedback to the correct source. With one click, the Contact Us tool can send an email to an institution's helpdesk, with all relevant info automatically included.
Email	mailbox	Send mail to participants in your site using the institution's official email system.
Email archive	mailarchive	Create a mailing list for your course or worksite, allowing for many-to-many broadcasting; access and search an archive of emails sent to this address.
Forums	forums	Create, moderate and manage discussion topics and groups within a course and send private messages to participants; options to grade discussions, set word count, permissions, etc.
Messages	messages	Send, receive and reply to messages among site participants using Sakai's internal email messaging.
Sign-Up	signup	Create and manage electronic sign-up sheets for events or activities such as office hours, presentation slots, research topics, review sessions, etc.
Wiki	rwiki	Create, share and edit web content collaboratively.

Grading and Assessment

Tool	ID	Description
Assignments	assignment	Create and grade online or offline assignments with a variety of options and settings, including honour pledge, peer review, and group submissions. Document preview lets instructors view and provide feedback in the browser without downloading student submissions.
Gradebook	gradebookng	Calculate, store, and distribute grade information to students. It uses a spreadsheet-style interface and is designed to provide faster, easier grade-entry capabilities.
Post'em	postem	Post, store, and distribute scores or non-numeric feedback and comments to students.
Rubrics	rubrics	Create, manage, and share grading rubrics that work within Sakai's Assignments, Forums, Gradebook, and Tests & Quizzes tools to provide rich feedback and automatically tabulated scoring.
Tests & Quizzes	samigo	Create and manage online assessments using a robust testing engine which includes a variety of question types, question pools, statistics, security options for high stakes testing, delivery time and date exceptions, and more.
Questions and Answers	qna	This tool allows members of a site to interact with each other by privately or publicly asking questions and receiving answers.

Course Development and Delivery

Tool	ID	Description
Cloud storage integrations	-	OneDrive and Google Drive are available in the Sakai file picker wherever a user attaches files.
Drop box tool	dropbox	Share files privately with course, project, or site participants.
External tool	basicti_unconfigured	Configure and launch external tools using IMS Learning Tools Interoperability.
Lessons	lessonbuildertool	Create and organise text, resources, quizzes, tests, assignments, links, videos, and other media into lessons or units. Lessons can be configured for conditional release of content, student-authored pages, group collaboration, and much more.
News	simple.rss	Display custom news content from dynamic online sources via RSS.
Overview	overview	View the site description and recent announcements, messages, discussion posts, and calendar events on the Overview landing page.
Podcasts	podcasts	Manage individual podcasts and podcast feed information for your site.
Resources	resources	Post, store, and organise material related to the course or project site. Drag and drop and WebDAV functionality make bulk file uploads quick and easy.
Search	search	Search content in a site using Sakai's Elasticsearch powered search tool.
Syllabus	syllabus	Post a summary outline of course requirements.
Web content	web	Display external web pages as part of the site menu.

Powerful System Administration

Course Management

Tool	ID	Description
Polls	poll	Easily create quick polls to collect and distribute data from site participants.
Roster	roster	View a list of course or project participants along with their pictures and profiles.
Section Info	sections	Manage multiple sections and assign teaching assistants to specific sections within a course.
Site Info	groupmanager	View and update information about the site, including site settings, tools, membership, groups, content imports, and more.
Statistics	sitestats	Use summary and custom reports to display site statistics regarding user visits, tool use, and course activity.

User Tools

Tool	ID	Description
Account		View and update user account details.
Help	help	Access context-sensitive help articles throughout the system.
Membership	site.membership	View and search for your enrolled sites or self-enrol in a joinable site.
New Account	admin_user	Create a new user account in the system.
Preferences	prefs	Set your individual preferences for notifications, time zone, language, active sites, and rich text editor mode.
Profile	profile	Create a personal profile and connect with other users in the system. Users can even record their name pronunciation as part of their profiles.
Reset Password	user	Reset your Sakai password using the self-service reset option.
Worksite Setup	siteinfo	View information about your sites, and (for users with appropriate permissions) create new sites and make changes to site settings.

System Administrator Tools

Tool	ID	Description
Admin Site Permissions	-	Quickly and easily add or remove permissions across all sites for specified roles.
Aliases	useralias	View, create, and edit alias strings for target items such as site email list addresses, friendly site URLs, etc.
Become User	su	Log in as another user to perform troubleshooting or diagnose issues without needing to know their password.
Delegated Access	-	Distribute admin functions or course access for specific users by departments or organisational units.
Email Templates	-	Create customised email notifications for users in your local Sakai instance to replace default system messages.
External Tools	basicti , basicti_unconfigured	Configure LTI Apps at the system level for use in all sites or just specific courses as needed.
Job Scheduler	schedule	Create simple or complex schedules for executing Quartz jobs, from the smallest stand-alone application to the largest e-commerce system.
Memory	-	View system cache sizes in order to better optimise performance.
Message Bundle Manager	-	Locate and modify system message properties, which define the user interface text for Sakai components in all supported languages.
MOTD or Announcements Online	-	Post system-wide announcements using the “Message of the Day” Announcement tool.
PA System	-	View information on currently running servers and sessions, as well as user IP address and browser information.
Realms	-	Deliver customisable system-wide or targeted alerts to Sakai users as either popup alerts or text banners, colour-coded by alert priority.
Sites	realm	View and customise roles and permissions for sites and site templates.
Site Archive	site	Back up sites or import existing backups individually or in bulk.
Sitestats Admin	sitestats	Create, modify, delete, and search for course, project, and user sites.
Tags Service	-	Display server-wide reports and save custom reports, which are available for instructors within all sites in the Statistics tool.
User Membership	-	Create and maintain tag collections either manually or by importing collections. Tags can be used in question pools and other areas of Sakai for search purposes.
Users	user	Locate, view, and modify site and group membership information for Sakai users in the system.
		Create, search, view, and update Sakai user account information.

Community Contributed Tools

The community Contributed tools are supported by various universities and are available in their own GitHub profile “sakaicontrib” (Apereo, 2023a).

Tool	ID	Description
Attendance	attendance	Take attendance online, print sign-in sheets, and assign attendance grades.
Blogwow	blogwow	Post blog entries which can either be viewed by the rest of the site members or kept private by the author. (Alternative to the Clog tool.)
Certification	-	Award printable certificates to site participants based on successful demonstration of specific criteria.

Clog	-	Post blog entries which can either be viewed by the rest of the site members or kept private by the author. (Alternative to the BlogWow tool.)
Evaluation System	evaluation	Create and deliver formative and summative (end-of-term) evaluations and surveys of Sakai course sites individually, hierarchically, or system-wide.
Mneme	mneme	Create and deliver tests and surveys in Sakai. (Alternative to the Tests & Quizzes core tool.)

External App Integrations

Tool	ID	Description
BigBlueButton	web_bigbluebutton	BigBlueButton is a purpose-built virtual classroom that empowers teachers to teach and learners to learn.
Cengage	web_cengage_owl	Developed by teaching chemists, OWLv2 is the online homework platform that helps students learn and master concepts, offering just-in-time support, study resources and eTextbook access.
Digication	-	Digication's learning platform is an integrative teaching, learning, and assessment software for students to showcase and share their learning.
Examity	web_examity	Examity is the leading live proctoring expert.
Explorance Blue	lti_explorance_blue	Explorance Blue is a mobile-friendly software that evaluates learning with formative and summative assessments. Increased accessibility and participation
Gradescope	gradescope	Gradescope helps you seamlessly administer and grade all of your assessments, whether online or in-class.
Hypothesis	web_hypothesis	Hypothesis is an easy-to-use pedagogical tool that enables students and teachers to have conversations in the margins of digital texts.
iClicker	-	iClicker is the market leader in student and audience response systems for Higher Education, recognised for ease of use, reliability, and focus on pedagogy.
iRubric	-	iRubric is a comprehensive rubric development, assessment, and sharing tool. Designed from the ground up, iRubric supports a variety of applications in an easy-to-use package.
Kaltura	-	Kaltura is driving communication, learning, and TV experiences for millions of users daily. Visit to learn more.
Karuta	lti_karuta	Karuta is a flexible tool for incremental prototyping and diffusion on the web of digital portfolios or ePortfolios for various purposes: showcase portfolio, learning portfolio, and assessment portfolio.
McGraw-Hill (mhcampus)	web_- mcgraw_hill_connect	McGraw-Hill Campus is a groundbreaking service that puts world-class digital learning resources just a click away for your faculty and students.
Office 365 Opencast	opencast_series	Microsoft 365 is our cloud-powered productivity platform. Opencast is a flexible, reliable, and scalable open-source video management system for academic institutions, built by a community of developers from leading universities and organisations worldwide.
OnTask	web_ontask	The OnTask Project aims to improve the academic experience of students through the delivery of timely, personalised, and actionable student feedback throughout their participation in a course.
Panopto	-	Panopto is the leading video platform for businesses and universities.
Perusall	-	With Perusall, an online social annotation platform, you can increase student engagement, collaboration, and community within your course.

Tsugi App Store	opencast_setup, basiclti_mecmovies, basiclti_studentnotes	Tsugi Cloud. An App Store for Education The tools hosted on this site can be seamlessly integrated into any Learning Management System that supports IMS LTI.
Turnitin	-	Turnitin solutions promote academic integrity, streamline grading and feedback, deter plagiarism, and improve student outcomes.
Webassign	web_webassign	WebAssign is the leading provider of powerful online instructional tools for faculty and students. In brief, instructors create assignments online within WebAssign and electronically transmit them to their class. Students enter their answers online, and WebAssign automatically grades the assignment and gives students instant feedback on their performance.
Zoom	web_zoom	Modernise workflows with Zoom's trusted collaboration tools.

Appendix – LMS Emotional Footprint

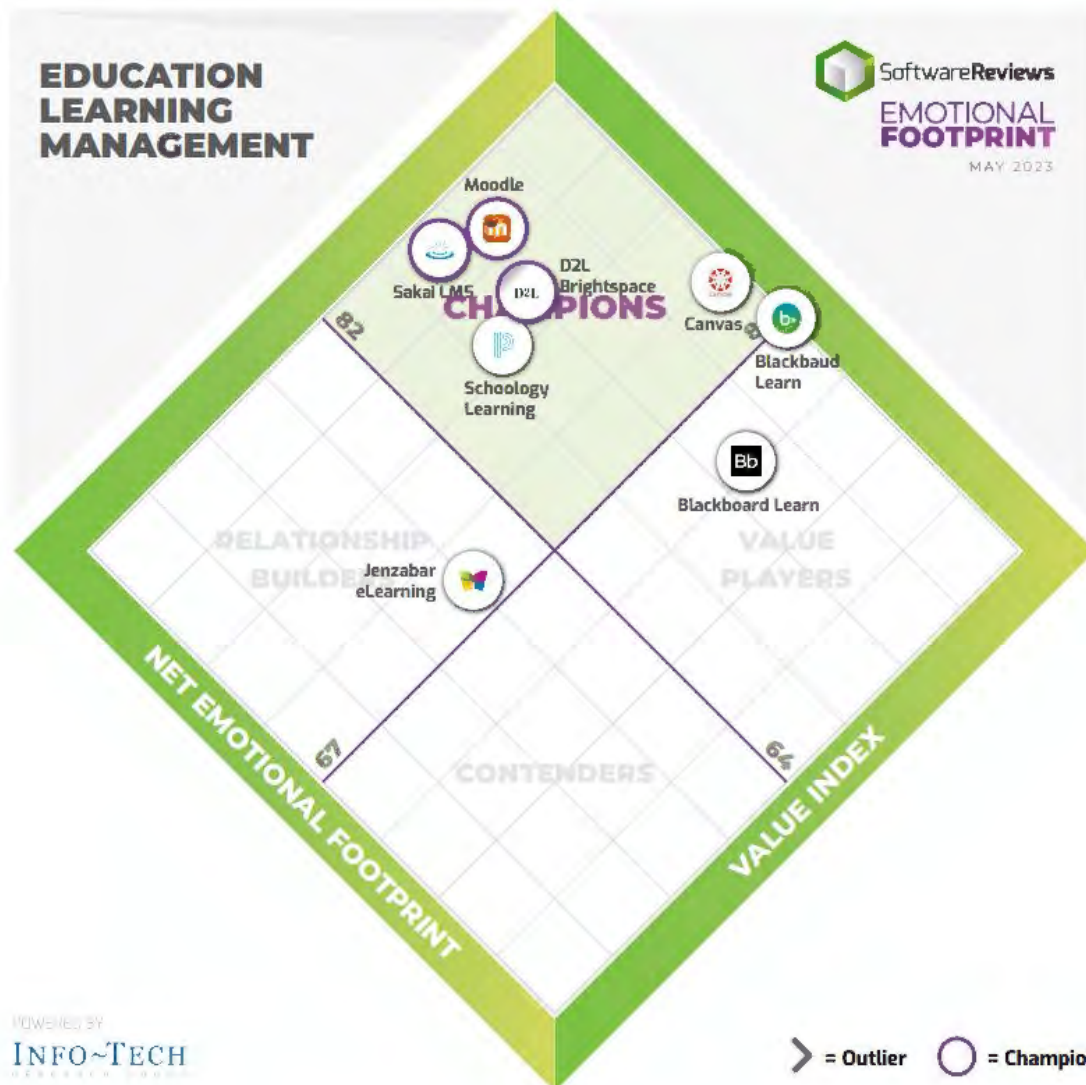


Figure 89: Software Reviews - Emotional Footprint for Education Learning Management System (SoftwareReviews, 2020)

Category Overview

This page provides a high level summary of product performance within the Education Learning Management category. Products are ranked by a composite satisfaction score (Composite Score) that averages four different areas of evaluation: Net Emotional Footprint, Vendor Capabilities, Product Features, and Likeliness to Recommend. The Net Emotional Footprint Score measures user emotional response ratings of the vendor (e.g. trustworthy, respectful, fair).

Use this data to get a sense of the field, and to see how the products you're considering stack up.

RANK	PRODUCT	COMPOSITE SCORE	NET EMOTIONAL FOOTPRINT	NET EMOTIONAL FOOTPRINT DISTRIBUTION	VENDOR CAPABILITIES	PRODUCT FEATURES	LIKELINESS TO RECOMMEND	NUMBER OF REVIEWS
	Sakai LMS	8.5/10	+90	2% NEGATIVE 92% POSITIVE	81%	83%	87%	98
	Moodle LMS	8.0/10	+81	4% NEGATIVE 85% POSITIVE	76%	79%	82%	206
	Canvas	7.9/10	+77	6% NEGATIVE 83% POSITIVE	77%	79%	84%	197
4	Schoology Learning	7.5/10	+76	6% NEGATIVE 82% POSITIVE	73%	74%	78%	106
5	D2L D2L Brightspace	7.5/10	+76	5% NEGATIVE 79% POSITIVE	76%	77%	72%	51
6	Blackbaud Learn	7.4/10	+69	8% NEGATIVE 77% POSITIVE	75%	74%	79%	42
7	Blackboard LMS Platform	7.4/10	+66	9% NEGATIVE 75% POSITIVE	75%	77%	76%	305
8	Jenzabar eLearning	7.0/10	+73	7% NEGATIVE 80% POSITIVE	71%	75%	62%	34

Figure 90: SoftwareReviews - LMS Category overview of product performance (Complete)
(Best Education Learning Management (ELMS) Systems 2023, 2023)

Vendor Capability Summary

This page summarizes user satisfaction with a variety of vendor capabilities regarding their product offering(s). Look for strong and consistent performance across the board when assembling your shortlist, and follow-up on areas of concern during the evaluation and negotiation processes

PRODUCT	OVERALL CAPABILITY SATISFACTION	BUSINESS VALUE CREATED	BREADTH OF FEATURES	QUALITY OF FEATURES	PRODUCT STRATEGY AND RATE OF IMPROVEMENT	USABILITY AND INTUITIVENESS	VENDOR SUPPORT	EASE OF DATA INTEGRATION	EASE OF ADMINISTRATION	EASE OF CUSTOMIZATION	AVAILABILITY AND QUALITY OF TRAINING	EASE OF IMPLEMENTATION
Sakai	80%	83%	85%	78%	80%	79%	82%	81%	79%	81%	73%	79%
Canvas	77%	79%	79%	79%	74%	77%	75%	76%	76%	74%	73%	80%
D2L Brightspace	76%	74%	76%	79%	72%	75%	73%	77%	79%	74%	78%	78%
Moodle LMS	76%	76%	77%	78%	75%	74%	74%	75%	76%	76%	74%	77%
Blackbaud Learn	76%	75%	79%	78%	75%	76%	74%	71%	81%	65%	78%	79%
Blackboard LMS Platform	75%	78%	78%	75%	74%	74%	72%	78%	76%	72%	71%	76%
Schoology Learning	73%	73%	73%	73%	69%	78%	74%	73%	75%	71%	72%	74%
Jenzabar eLearning	71%	72%	67%	67%	69%	75%	73%	71%	74%	75%	63%	71%
CATEGORY AVERAGE	76%	76%	77%	76%	73%	76%	75%	75%	77%	74%	73%	77%

Figure 91: SoftwareReviews - LMS Category overview of Vendor Capabilities (Best Education Learning Management (ELMS) Systems 2023, 2023)

Appendix – 2019 Academic Year

Source: <https://www.studentroom.co.za/uct-term-dates-2018-2019-what-are-the-important-dates-for-university-of-cape-town/>

	2019	Duration
1ST QUARTER	11 February – 12 April Week 7 - 15	8 weeks 4 days, 60 days, 44 weekdays
1ST VACATION	13 April – 22 April Week 15 - 17	1 week 2 days, 9 days, 5 weekdays
2ND QUARTER	23 April – 09 June Week 17 - 23	6 weeks 5 days, 47 days, 33 weekdays
MID-YEAR VACATION	10 June – 14 July Week 24 - 28	4 weeks 6 days, 34 days, 24 weekdays
3RD QUARTER	15 July – 23 August Week 29 - 34	5 weeks 4 days, 39 days, 29 weekdays
MID-TERM VACATION	24 August – 01 September Week 34 - 35	1 week 1 day, 8 days, 5 weekdays
4TH QUARTER	02 September – 24 December Week 35 - 52	16 weeks 1 day, 113 days, 81 weekdays

Table 68: UCT 2019 Academic Year

Appendix – Participant Data Set - DIANA

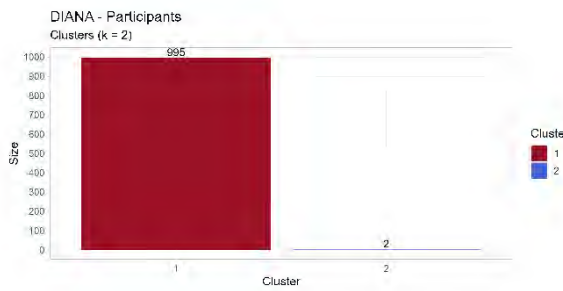


Figure 92: DIANA (k=2) size of participant clusters



Figure 93: DIANA (k=2) cluster silhouette of participants

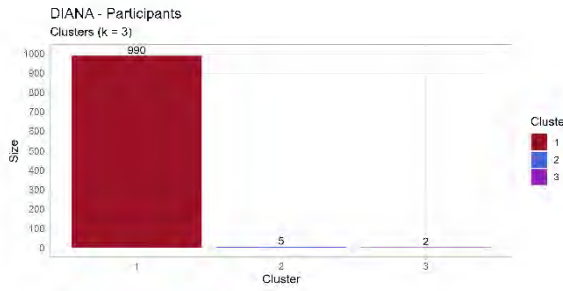


Figure 94: DIANA (k=3) size of participant clusters



Figure 95: DIANA (k=3) cluster silhouette of participants

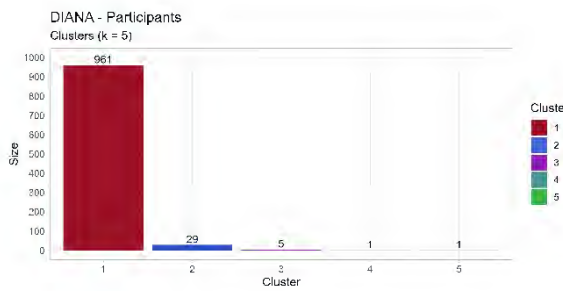


Figure 96: DIANA (k=5) size of participant clusters

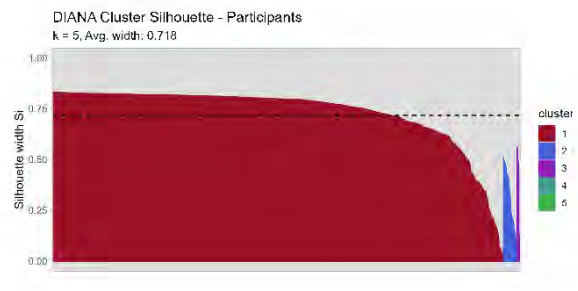


Figure 97: DIANA (k=5) cluster silhouette of participants

Correlation Between Tools Visible to Students
(Categories, n=997)

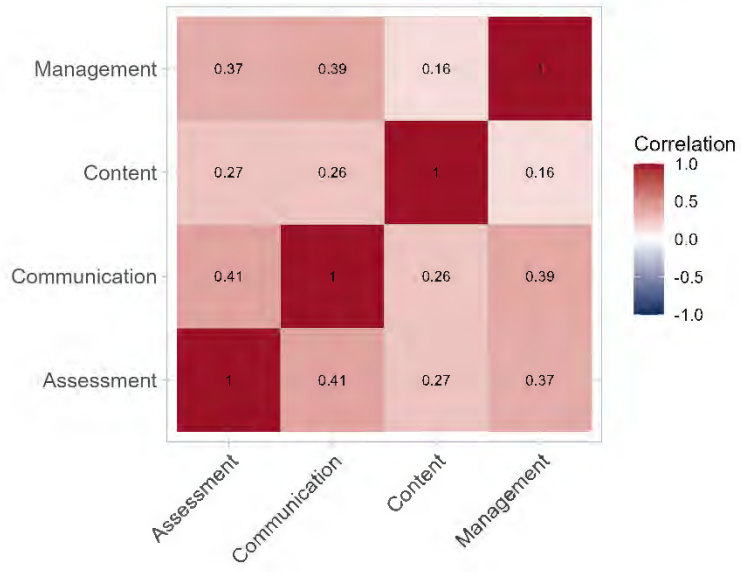


Figure 99: Correlation of tool categories in “Tools Visible to Students”

Appendix – Tools Visible to Students - PCA

There are 42 different tools visible to students in this data set of 997 course sites. A PCA is run to reduce the variables and improve interpretability, and the appropriate cut-off value (90%) decides the number of dimensions to use. In this case, that is 32 dimensions, as shown on the next page, with the eigenvalues of each dimension and the cumulative variance percentage. Choosing 32 dimensions reduces the variables by 23.81%. The first dimension in the PCA describes 3.533 variables and 8.42% of the variance.

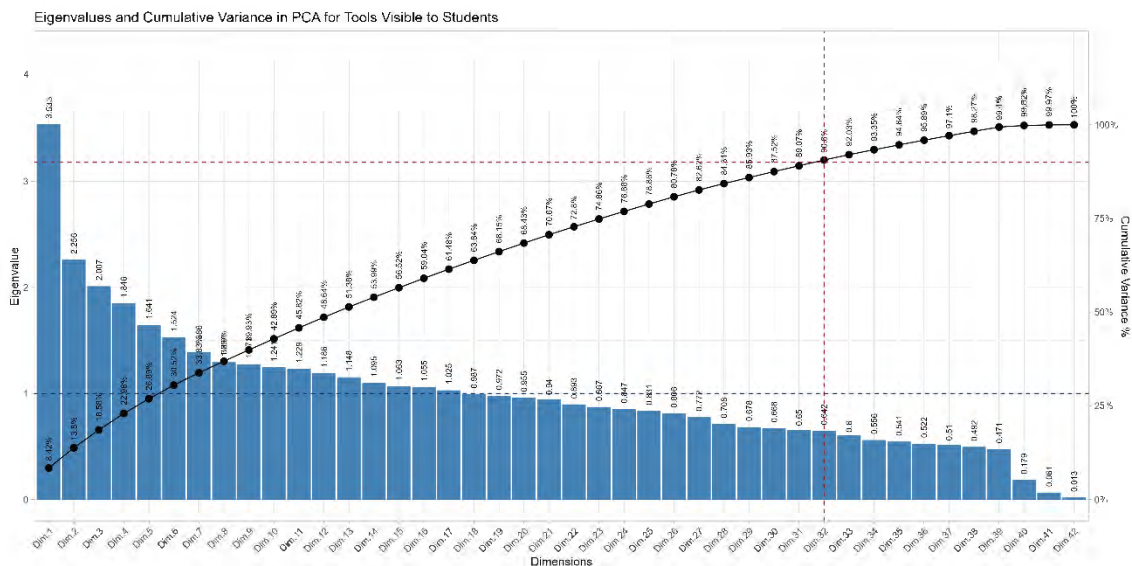


Table 70: Tools visible to students – PCA – Eigenvalues and Cumulative Variance

The next page contains the first six components of the PCA run for tools visible to students and the biplot for the first eight components. The biplot vectors are coloured using the cos2 score for the contribution value for each vector. The size of the vectors indicates how strongly they influence the principal components and their direction in the relationship with the other variables. If vectors (arrows) point in the same direction (small angle between them), the variables they represent are positively correlated. Otherwise, when the vectors diverge (pointing in opposite directions), those components are negatively correlated.

The table shows that the communication and collaboration tools like calendar, schedule, messagecenter, and forum each contribute close to 12% to the first principal component, also visible in the biplot by the blue colour for those vectors (Dim-1 ~ Dim-2). The second principal component is calculated with the condition that its components are perpendicular (uncorrelated) to the first principal component, accounting for the next highest variance. Dim-2 shows that webassign, eBook LTI and assignment contribute mostly to its variance.

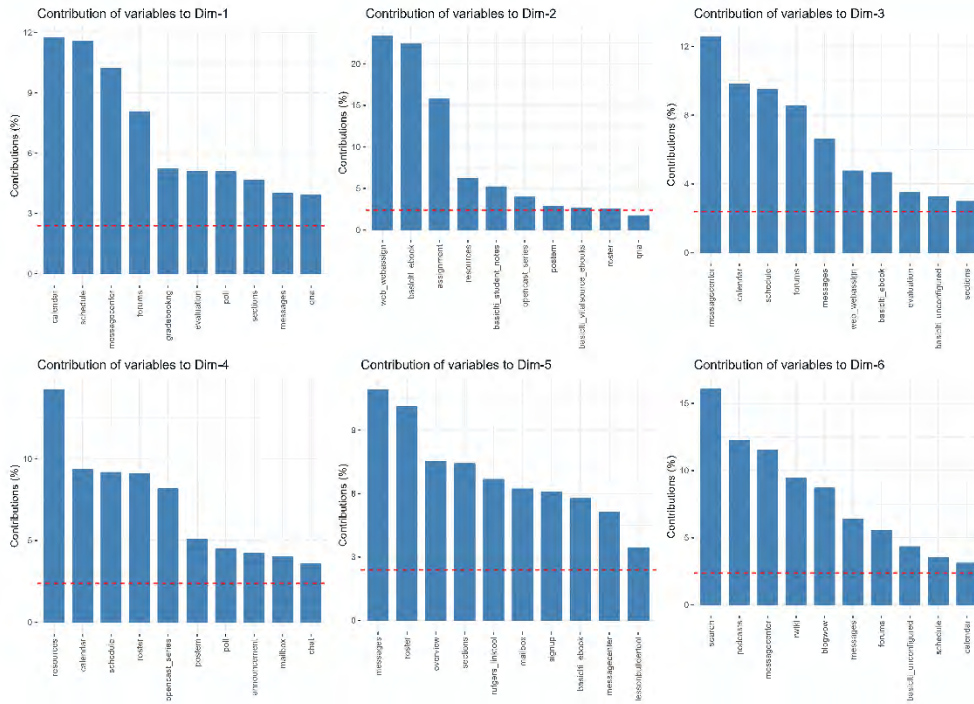


Figure 100: Contribution of variables to the PCA components for tools visible to students

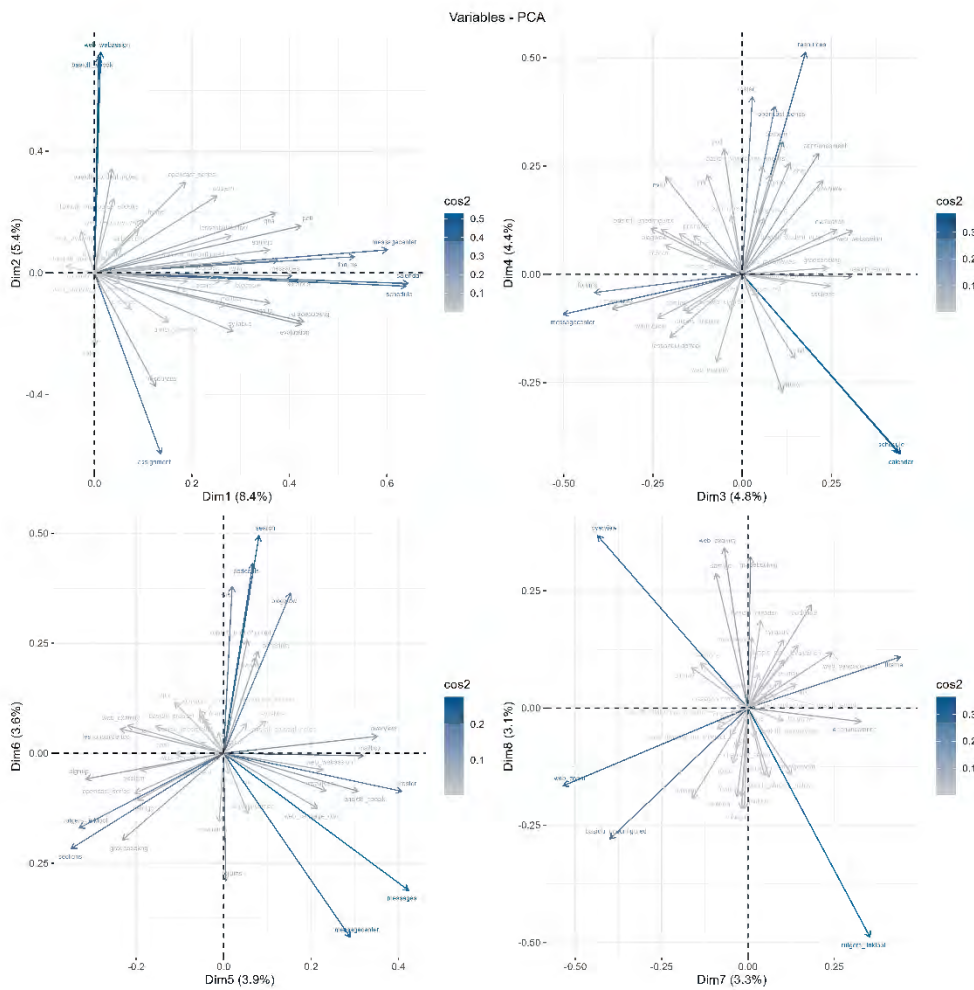


Figure 101: Biplot of principal components, including Cos2 contribution values for PCA

These vectors also meet the calendar and schedule at just over 90°, indicating they are not correlated. This can be expected from the relationship variables in the first and second principal components. The fourth principal component introduces the content resources tool, which is negatively correlated with the schedule and calendar tools, which are also high percentage contributors to this dimension. It's interesting to note that the iframe tool is not correlated with the Rutgers link tool, even though they perform a similar task.

The values of the biplot can be represented in a cos2 correlation graph. The first ten components are shown below. This table visually represents the contribution of the variables to the principal components.

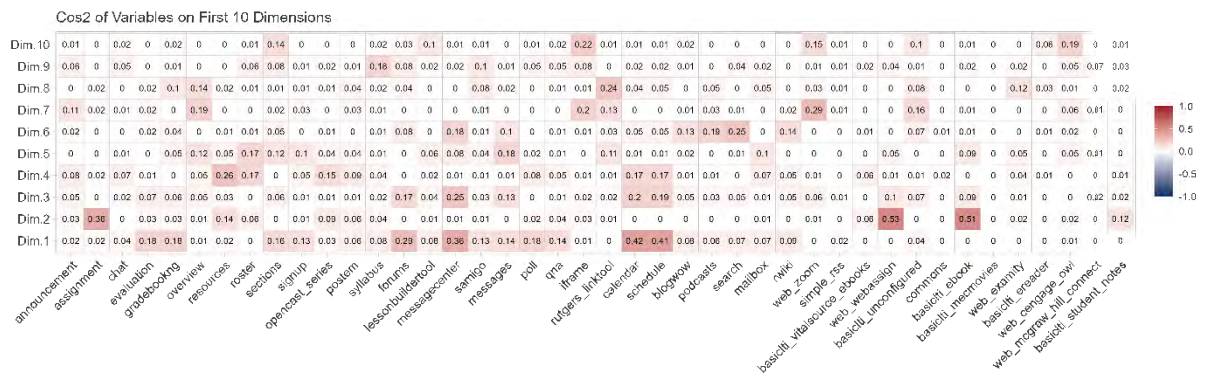


Figure 102: PCA – Cos2 of variables on the first 10 dimensions in tools visible to students